DISSERTATION

MATHEMATICAL AND EXPERIMENTAL STUDIES IN CELLULAR DECISION MAKING

Submitted by

Samanthe Merrick Lyons

Graduate Degree Program in Bioengineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Ashok Prasad

June Medford
Chris Snow
John Kisiday

ABSTRACT

MATHEMATICAL AND EXPERIMENTAL STUDIES IN CELLULAR DECISION MAKING

The biological sciences are undergoing an epistemological revolution. Mathematical modeling, quantitative experiments and data analysis, machine learning and other methods of “big-data” modeling are slowly but surely changing the way the biological and biomedical sciences and engineering are being carried out. This thesis presents work that seeks to advance understanding of biological processes using mathematical modeling as well as experiments coupled with sophisticated quantitative analysis. The central theme of the research presented is cellular decision-making. A cellular decision is defined here as a transition from one cell state, or phenotype, to another, based upon information received from an external or internal signal.

This work explores the mechanisms behind cellular decisions with three specific systems and a variety of mathematical and modeling techniques. This dissertation begins with a brief survey of the use of mathematical modeling in cellular biology, utilizing specific example of various approaches. This reviews the diversity of techniques available from detailed mechanistic models to simplified phenomenological representations, and notes some applications demonstrating the utility of such models.

The first exploration of cellular decisions is concerned with the question of how cells can make decisions in the face of cross-talk from multiple signals. The real cellular environment is noisy, with stochastically varying levels of external signals and cellular decisions required in spite of this noise. In Chapter 3 the ubiquitous bacterial two-component signaling system and the similarly structured mammalian TGF-β pathway are modeled with stochastic simulations of the

chemical master equation. Information theory is utilized to quantify the amount of information transmitted by these signaling systems in the presence of competing signals from cross-talk, revealing that the mammalian TGF- pathway was able to transmit information accurately despite high levels of cross-talk, while the bacterial two-component system, due to a smaller system size and the structure of phospho-transfer rather than phospho-relay, was poor at discriminating from competing cross-talk. This work presents a novel thesis: many signal transduction systems suffer less from cross-talk than was commonly imagined, and may actually make use of cross-talk for cross-regulation.

The second system of cellular decisions studied in this work is a bistable synthetic toggle switch network motif composed of mutually repressible promoters in Chapter 4. This motif has been widely studied in isolation for its dynamical and static properties. However, the behavior of these switches has never previously been analyzed when coupled with a downstream binding partner, termed a “load”. Real toggle switches, whether synthetic or natural, always have loads connected with them. The toggle-switch system was modeled mathematically with ordinary differential equations as well as using stochastic simulations of the chemical master equation to determine the effect of a load. The quasi-potential energy landscape of the bistable switch was calculated utilizing a novel method which revealed that, in some parameter spaces, a downstream component can significantly alter the stability of the switch; addition of a positive feedback loop could provide for a tunable switch.

Chapter 5 is concerned with developing methods for identifying a complex cellular transition from less metastatic to more metastatic cancer cells. The importance of metastatic disease in the pathology of cancer cannot be understated as it is the cause of 90% of deaths from cancer. The process by which cancerous cells become metastatic is complex, but requires

specific cellular mechanical conditions in order to occur. The use of cancer cell shape to predict metastatic behavior in pathology samples is a key component of prognostication, however *in vitro* cancer cell shape is less commonly studied. This work developed a mathematical algorithm to extract shape parameters from images of cancer cells and applied multiple statistical techniques to elucidate differences between metastatic and non-metastatic cancer cells. While both simple and complex statistical techniques including t-tests, principle component analysis (PCA) and non-metric multidimensional scaling (NMDS) revealed distinct changes, the population of cells from highly metastatic and less metastatic paired osteosarcoma cells showed significant overlap. Machine learning algorithms were, however, able to successfully classify samples of cells to high or low metastatic lines with high accuracy.

The concluding chapter presents a brief analysis of the new questions that this research has elucidated, and delineates some future tasks to address them.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Ashok Prasad for his support and guidance throughout my time at CSU. I would like to express my utmost gratitude to those who contributed to my work and helped me to become the researcher I am today, especially to Elaheh Alizade, Dr. June Medford and those in the Medford lab, and all co-authors on my publications. Thank you to all current and past members of the Prasad lab for your comradery and support. Thank you also to my committee – Drs. Medford, Kisiday and Snow – for your time and guidance.

Thank you to my fellow veterinary classmates and to my veterinary professors for your support during the pursuit of a dual degree, especially Dr. Olivia Love, without whom I could not have completed both degrees simultaneously. Thank you to Dr. Erik Johnson for your consultations and humor. Thank you to Dr. Chuck Miller for inspiration to pursue a career in veterinary medicine and for modeling the type of veterinarian I aspire to be. Thank you to the Bear Valley Veterinary Care Center and Dr. Carola Stevenson for the flexibility to complete this dissertation while working as a veterinarian and support as a new veterinary graduate.

Lastly, I am grateful for my family and friends, both near and far, for their support and encouragement. Especially to my parents, siblings, Cat, David and Roxy.

TABLE OF CONTENTS

CHAPTER 1: COMPUTATIONAL MODELING IN ANALYSIS OF CELLULAR DECISION MAKING

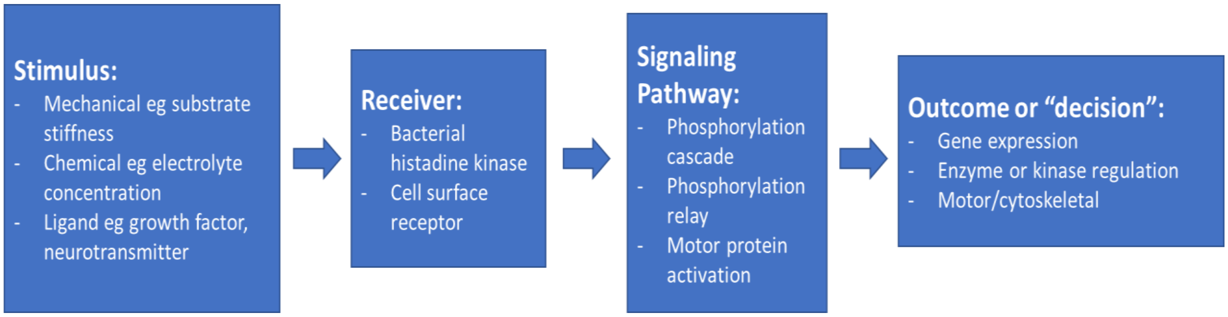
1.1 Cellular Decision Making

Cells are complex, with thousands of coding genes, proteins, and functions, all connected in intricate networks resulting in seemingly elaborate decisions. A variety of cellular phenomena from apoptosis, division, differentiation, chemotaxis and metastasis can be thought of as cellular decisions made as a result of some stimulus about the external world or an intracellular state [1]. A cellular decision in this context is a transition from one cellular state – often considered phenotype – to another, consequent upon receiving external (such as cell surface stiffness or growth factors) or internal (such as damaged DNA) signals. A central theme of this work has been the study of protein circuits that underlie cellular decision-making in different contexts and the utilization of computational modeling techniques to advance understanding of these decision-making processes.

Cells make many simple and complex decisions daily, whether it is a neutrophil orienting to a bacterium through chemotaxis [2]; a stem cell differentiating into an osteoblast during development or fracture healing [3]; or a cancer cell undergoing epithelial-to-mesenchymal transition and distant metastasis [4] these decisions have significant biological importance.

1.1.1 Simple Decisions

In a simple model of cellular-decision making, a stimulus is conveyed to the cell through a signal transduction network concluding in a change. Often this decision is manifest through a change in gene expression profile although faster changes may be brought about by regulation of enzymes, regulation of motor activity or changes in cytoskeletal properties, or changes in kinase activity prompting further signaling pathway activity.



Stimulus:
- Mechanical eg substrate stiffness
- Chemical eg electrolyte concentration
- Ligand eg growth factor, neurotransmitter

Receiver:
- Bacterial histadine kinase
- Cell surface receptor

Signaling Pathway:
- Phosphorylation cascade
- Phosphorylation relay
- Motor protein activation

Outcome or "decision":
- Gene expression
- Enzyme or kinase regulation
- Motor/cytoskeletal

Figure 1.1 Schematic of Cellular Signal Transduction Network

Examples of simple decisions are found throughout biology; bacterial two-component signaling pathways are one of the most ubiquitous examples [5], [6]. These simple pathways regulate a myriad of cellular decisions from nutrient sensing, chemotaxis, osmolality control, quorum sensing and many others. One of the best characterized is bacterial quorum sensing [7] where an external signal produced by neighboring cells binds a surface receptor, causing a signaling pathway that results in changes in gene expression – such as sporulation, biofilm formation, competence, conjugation and others. Some properties of this two-component system and a similarly structured mammalian signaling pathway – TGF-β – are explored in Chapter 3, Cross-Talk and Information Transfer in Mammalian and Bacterial Signaling. The TGF-β family of growth factors is responsible for decisions of mesenchymal stem cells differentiating into bone, for example [8]. Simple decisions can also be observed in plants; some angiosperm plants which utilize sexual reproduction 'decide' to accept 'non-self' pollen but reject 'self' pollen via triggering programmed cell death, thus preventing in-breeding [9].

1.1.2 Information Processing in Decision Making

Many cellular choices can be broken into a digital or binary output, a decision between a few discrete options – such as divide or not, move or stay put, differentiate or stay the same, option A or B. One well characterized example is the quorum sensing pathway of *V. fischeri*, a bacterium which synergistically lives in eukaryotic cells and produces light [10]. At low cell populations, the LuxI protein, stimulates gene production of an autoinducer. This autoinducer builds up in the environment and at high cell populations, the autoinducer activates the LuxR protein which in turn activates an operon to produce light. The "decision" to produce luciferase therefore requires transmitting the signal of cell population, via the concentration of autoinducer, to the cell nucleus. The cell integrates the concentration of the LuxI and subsequent LuxR proteins to determine if cell population is low – and not produce light – or if cell population is high and produce light.

Decision-making is therefore inherently tied to information transfer and processing. To make the appropriate decision, a cell must be able to interpret the information given by these stimuli – in this example the autoinducer – and transduced through a signaling pathway. Efficient information transduction is required for accurate signaling, and in turn accurate decision-making.

Information theory, a mathematical framework that quantifies the efficiency of information transferred through a noisy channel, such as a cellular signaling pathway, provides an ideal framework to understand the interdependence of information transfer and cellular decision making [11]. This framework is utilized to quantify information transfer efficiency in two similar signaling pathways in Chapter 3, Cross-Talk and Information Transfer in Mammalian and Bacterial Signaling.

1.1.3 Synthetic Decision Making

A rather new field of biology is synthetic biology, which is a rapidly growing field that utilizes biology and engineering to design novel biological components, systems and functions [12]. This field applies computer science to biology through the use of advanced genetic engineering with the objective of developing complex artificial systems such as large-scale production of an important antimalarial drug and synthetic circuits that allow for coordinated expansion of biomass to produce fuel such as ethanol [13].

The key characteristic that distinguishes synthetic biology from genetic engineering is its emphasis on the engineering of protein and RNA circuits for controlling an aspect of cell behavior and imparting a novel functionality to the host organism. Synthetic biology therefore has concentrated on construction of simple decision circuits that perform information processing. The most notable examples here are the genetic toggle switch and logic gates [14].

In some forms of cellular decision making, a cell must convert an analog signal – concentrations of proteins, molecules, and transcription factors – into a digital or binary output. There are many genetic and signaling motifs that may result in this conversion. Some are ubiquitous in nature, such as the bacterial two-component system, while others are created by synthetic biologists designing novel gene transcription and protein signaling networks [15]. Synthetic biology allows us to study these decision-making processes in isolation which can provide a deep understanding of the basic science behind a cellular decision.

The basic protein network or “circuit” behind digital decisions, or switches, in cells are known and have been studied previously using modeling, experiments and synthetic construction. However, what has been generally ignored is that these circuits, to function effectively, have to be linked with the downstream network. This work explores the consequences of this link on biochemical switches using computational modeling based on both ordinary differential equations (ODEs) and stochastic simulations in chapter 4.

1.2 Complex Decision Making

In addition to the simple cellular decisions based on one or two data inputs, cells also make complex decisions requiring analysis of multiple pieces of information about the cell and its environment. Two complex, well-studied, and extremely important examples of complex cellular decisions are apoptosis and metastasis; in this context, a “decision” is a change in cellular state arising from a cellular network synthesizing information from an internal or external signal. Complex decisions are harder to study mechanistically as they involve a number of processes that are linked together; however, they are very important for us to identify, because of their functional importance in cellular life. Here we briefly discuss cellular apoptosis and metastasis as examples of complex decision. This thesis develops a hypothesis and methods for the detection of complex cellular decisions using shape changes, applied in particular to metastatic cancer cells.

1.2.1 Cellular apoptosis

Cellular apoptosis, or programmed cell death, may be playfully conveyed with Shakespeare’s infamous line: “to be, or not to be: that is the question.” This process is crucial to the successful functioning of many aspects of cellular biology including normal cell turnover, immune system function, embryonic development and differentiation [16]. Failure of apoptosis to occur at the correct rate and time is implicated in autoimmune disease, dysfunctional development, Alzheimer’s and Parkinson’s diseases, and is an essential component to cancer.

Such an important mechanism should be tightly regulated and there are multiple checks in the process to trigger cell death appropriately. One well-characterized route for programmed cell death features a complex network of positive and negative feedback loops in the p53-dependent pathway, which is triggered based upon signals from damaged DNA such as when a cell is exposed to irradiation [17]. There are also many signals which prompt cell death from genetically programmed timing such as in multicellular organism cellular development [18], after exposure to harmful substances resulting in DNA damage [17], after exposure to heat, ischemic injury or hypoxia, or toxic compounds such as chemotherapy, or after identification as being damaged by immune system T-cells [19].

There are three primary apoptosis pathways: an extrinsic pathway through which apoptosis is triggered by death receptors and caspase 8 is activated, an intrinsic pathway through which mitochondrial changes occur following damage and caspase 9 is activated, and the perforin/granzyme pathway through which T cells activate caspase 3 and 10. All three pathways result in the activation of a singular 'execution' pathway via one common protein: caspase 3 [16]. The signaling pathways for each pathway are tightly regulated, but ultimately the numerous suppressing and activating inputs result in one binary decision: to be or not to be.

1.2.2 Neoplastic Cell Metastasis

Another example of a complex, binary cellular decision with immense importance is the 'decision' of a neoplastic cell to migrate out of the tissue of origin and invade other tissues. Metastasis, the development of secondary cancer growths at a distant site, is responsible for 90% of human deaths from cancer [20]. To form successful metastases, tumor cells must navigate a complex, multi-stage process including: detachment from primary tumor, migration to vascular supply, intravasation, survival and transit in blood or lymphatic vessels, extravasation, and successful growth and adhesion in a new site [21]. With such a complex, multistep process, a binary decision is an overly simplified way to view this process, however on a practical level, the binary outcome – the lack or presence of metastatic disease – is what ultimately has clinical significance. With metastatic state having such profound importance, much of clinical cancer diagnosis is focused on determining the likelihood and extent of metastasis through stage and grade. The stage of cancer assigns a number based on how large a tumor has grown and whether metastasis has occurred, and to where in the body. Grading assigns a number to the appearance of cancer cells based upon microscopic examination of shape, variability, and mitotic activity utilizing a set grading scheme based upon the type of cancer; this number has a predictive value in gauging future biologic behavior such as metastasis and growth.

The information from grade and stage are then interpreted to guide clinical decisions and treatment selection as well as prognosticate, in part based on likelihood of formation of metastasis. For example, canine cutaneous mast cell tumors (MCT), a common type of cancerous skin tumor in dogs, can be classified based on the Patniak grading scheme. This separates tumors into three grades based on histological appearance assessing the extent of involvement, cellularity, cell shape, mitotic index, and changes in normal tissue (stromal reaction) [22]. Dogs with the least aggressive, grade I tumors, can often be cured with surgical removal alone. In one study 93% of dogs with grade I MCTs were alive at 1500 days; by comparison, only 6% of dogs with the most aggressive, grade III, tumors were alive at 1500 days. Grade III MCT is associated with a high rate of metastasis (50-90%) whereas grade I rarely forms metastases [23]. In practice, a dog that has complete surgical removal of a grade I MCT does not often require follow-up while a dog with a grade III MCT is recommended to follow surgery with aggressive chemotherapy. Thus, grading in canine MCT has a significant prognostic value in predicting both metastatic disease as well as survival time; as a result, grading guides clinical decision making regarding additional diagnostics and treatments.

1.2.3 Epithelial-Mesenchymal Transition

Epithelial-mesenchymal transition (EMT) is the process of an epithelial cell becoming a mesenchymal cell by losing characteristics, such as cell-cell adhesion and cellular polarity, and by acquiring characteristics, such as migratory and invasive properties, resulting in multipotent stromal cells (MSCs) which are capable of differentiating into many different cell types [24]. This process occurs appropriately during embryogenesis and is necessary for development of the specialized epithelial and mesenchymal cell types [25], [26]. It is also involved in both successful tissue healing and regeneration and unsuccessful tissue healing resulting in organ fibrosis [27], [28]. Recent work has focused on EMT's role in cancer, particularly in tumor progression and the acquisition of characteristics which allow for metastasis, with the hypothesis that EMT plays a role in some cancer's acquisition of the ability to become more invasive locally and to form distant metastatic disease.[29]–[31].

EMT has been described as a cellular decision [24] as well as a model of metastasis [29], [30]. Of note, EMT occurs with distinct cellular shape changes of adoption of spindle-cell shape, increased motility, increased focal adhesion dynamics, loss of apical-base polarity, and loss of cell-cell and cell-basement membrane adhesions; the resulting mesenchymal cells have a dramatically flattened, elongated leading-trailing morphology [32]–[36]. Some work has linked EMT and the resulting cellular shape changes with altered genetic expression of cytoskeletal proteins in metastatic cancer cells, providing a possible partial mechanism of the cellular shape changes and acquisition of metastatic phenotype [37]. The idea to predict cancer cell behavior and metastatic potential via assessment of *in vivo* cell shape via histopathology is not new [38]. The analysis of *in vitro* cultured cancer cell shape to assess metastatic potential is a newer application which has most commonly been used to probe mechanism of metastasis [39]–[41]. A less explored use of cell shape analysis of *in vitro* cultured cancer cells is assessment of metastatic potential. As demonstrated in the above described example of grading schemes of canine Mast Cell Tumors, the use of multiple sources of information regarding metastatic potential can greatly improve prognostication. The concept of supplementing cancer metastatic prognostication with *in vitro* cell shape paired with the opportunity to utilize analysis of cell shape to gain insights into the mechanism of acquisition of metastatic potential represented by shape changes motivated the work in Chapter 5: Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas. This work explores the use of morphological markers to identify complex cellular phenotype changes and applies advanced statistical analysis and machine learning models to make predictions of metastatic potential based upon cellular shape.

1.3 Modeling in Biology

The second unifying theme of this work is the utilization of computational modeling techniques to advance understanding of these cellular decision-making processes. Starting with simple models that can be mathematically simulated with exact stochastic simulations and differential equations this work concludes with large-scale machine learning models which predict metastatic potential of osteosarcoma cells and demonstrates the scope of biological modeling which can be applied to simple, synthetic, and complex cellular decisions. A review on the use of modeling in biology provides context to the rest of this thesis in Chapter 2.

1.4 Dissertation Statement

Chapter 1 introduces the topic and summarizes the chapters of this dissertation. Chapter 2 provides background information necessary to understanding the context of biological modeling and cellular decision-making. Chapter 3 applies information theory and stochastic simulation with the Gillespie algorithm to explore the effects of cross-talk on information transfer in mammalian and bacterial signaling pathways. Chapter 4 utilizes stochastic simulations of the toggle switch network motif to characterize the quasi-potential landscape of the toggle switch when it is connected to downstream components that apply a load. Chapter 5 details the use of a mathematical algorithm to extract shape parameters from images of cancer cells and multiple statistical techniques to elucidate differences between metastatic and non-metastatic cancer cells.

1.4.1 Chapter 2 – Mathematical modeling applications in cellular biology

The second chapter of this dissertation serves as an extended introduction that contextualizes the importance of computational modeling in biology in a review on "Mathematical modeling applications in cellular biology". This covers a brief summary of diverse applications of mathematical modeling in cellular biology and discusses the value of pairing mathematical models with experiment. Examples of the successes and limitations of utilizing modeling in biology and the immense value of modeling are reviewed. Objectives of various types of models, the relationship between models and experiment, and theory on how to develop a model are also briefly discussed.

1.4.2 Chapter 3 – Cross-Talk and Information Transfer in Mammalian and Bacterial Signaling

Information theory was applied to two similarly structured signaling pathways: the bacterial two-component system and the mammalian TGF-β pathway. The bacterial two-component signaling system is ubiquitous, utilizing phosphotransfer between a receptor and response regulator. The TGF-β pathway, activated by cell surface receptors binding TGF- β family ligands, signals via phosphorylation of SMAD proteins, resulting in many possible genetic changes. The multitude of possible inputs and similarity of intracellular transducers leads to the question of how well cells can distinguish between these signals. Using the Gillespie algorithm and stochastic simulations, both systems were modeled and information theory was used to address this question. The result was that with a single intracellular protein channel cells ability to discriminate between ligand input was poor; with two separate channels, discrimination ability was near perfect. Surprisingly, information transfer and ability discriminate between ligands are quite insensitive to high levels of cross-talk between the two signaling channels in the mammalian pathway but poor in the bacterial pathway. This difference was due to both pathway structure of phosphotransfer vs phosphorelay and system size with a smaller system size in smaller bacterial cells suffering robustness against cross-talk as a result. The suggestion that mammalian systems can tolerate high cross-talk may have played a role in the evolution of new functions by tolerance of small mutations resulting in cross-talk prior to evolutionary pressure to diverge into distinct channels. Conversely, the lack of observed cross-regulation in bacterial two-component systems may be in part due to loss of information in the presence of cross-talk.

1.4.3 Chapter 4 – Loads Bias Genetic and Signaling Switches in Synthetic and Natural Systems

A commonly utilized synthetic signaling motif – the toggle switch – was studied with the presence of interconnected ‘loads’ of other downstream signaling motifs in Chapter 4. Modularity can be a key assumption in synthetic biology, where it is important that, when network motifs are combined, they do not lose their essential characteristics; however, the interactions with downstream elements can result in changes of the dynamical equations describing upstream modules. This work stochastically simulated the toggle switch network motif and utilized a novel method to assess the potential energy landscape of a bistable switch. We demonstrated that connection to a downstream load does in fact affect function of the switch. By employing novel theoretical methods, we discovered that adding an additional downstream component to the simple toggle switch changes its dynamical properties by changing the underlying potential energy landscape. We also found that an additional motif found in naturally occurring toggle switches could tune the potential energy landscape in a desirable manner, providing a possible explanation for the existence of this additional regulatory protein in some natural toggle switches. This modeling work emphasizes the importance of incorporating effects of downstream components in modeling synthetic systems and design of networks.

1.4.4 Chapter 5 - Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas

The fifth chapter of this thesis applies advanced statistical analysis and machine learning methods to characterize and make predictions about the most complex cellular decision addressed in this work: the metastatic potential of cancer cells. Based upon the knowledge that metastatic cancer cells have altered cytoskeletal properties, shape characteristics of more metastatic cells are expected and observed to be distinct from their less-metastatic counterparts. The work assessed four paired lines of osteosarcoma where a highly-metastatic cell line was derived from a less-metastatic parental line through *in vivo* passage and selection. Statistical analysis of two-dimensional images of these cultured cells characterized morphological changes into two categories with the majority of cell lines demonstrating a more mesenchymal cellular morphology. A neural network algorithm was able to distinguish between the high- and low-metastatic cell lines with near-perfect accuracy, while other statistical methods were unable to distinguish between cell lines due to highly overlapping cellular morphology. This demonstrates a tight pairing between experiment and mathematical analysis in a collaborative project which combined cellular culture, imaging, development of algorithms to automate extraction of cell shape parameters in a high-throughput manner, multiple types of statistical analysis, and machine learning. As discussed in section 1.1.4, the ability to predict metastasis is hugely important in the diagnosis and treatment of cancer. This work developed an initial toolbox that works to extract large amounts of information from cellular experiments through quantifiable shape metrics. With refinement, this concept could provide personalized high-throughput data on cancer cell shape, and ultimately could be used to guide clinical decision making in conjunction with the gold standard of stage and grade, in addition to its use in gaining insight into mechanisms of metastasis.

REFERENCES

[1] G. Balázsi, A. van Oudenaarden, J. J. Collins, J. Sippy, M. Feiss, I. Golding, B. L. Bassler, N. P. Ong, M. C. Prevost, J. P. Latgé, and et al., "Cellular decision making and biological noise: from microbes to mammals.," *Cell*, vol. 144, no. 6, pp. 910–25, Mar. 2011.

[2] D. A. Bloes, D. Kretschmer, and A. Peschel, "Enemy attraction: bacterial agonists for leukocyte chemotaxis receptors," *Nat. Rev. Microbiol.*, vol. 13, no. 2, pp. 95–104, Dec. 2014.

[3] G. Chamberlain, J. Fox, B. Ashton, and J. Middleton, "Concise Review: Mesenchymal Stem Cells: Their Phenotype, Differentiation Capacity, Immunological Features, and Potential for Homing," *Stem Cells*, vol. 25, no. 11, pp. 2739–2749, Nov. 2007.

[4] J. Song, "EMT or apoptosis: a decision for TGF-β," *Cell Res.*, vol. 17, no. 4, pp. 289–290, Apr. 2007.

[5] V. L. Robinson, D. R. Buckler, and A. M. Stock, "A tale of two components: a novel kinase and a regulatory switch," *Nat. Struct. Biol.*, vol. 7, no. 8, pp. 626–633, Aug. 2000.

[6] A. M. Stock, V. L. Robinson, and P. N. Goudreau, "Two-Component Signal Transduction," *Annu. Rev. Biochem.*, vol. 69, no. 1, pp. 183–215, Jun. 2000.

[7] M. B. Miller and B. L. Bassler, "Quorum Sensing in Bacteria," *Annu. Rev. Microbiol.*, vol. 55, no. 1, pp. 165–199, Oct. 2001.

[8] A. H. Reddi, "Regulation of cartilage and bone differentiation by bone morphogenetic proteins.," *Curr. Opin. Cell Biol.*, vol. 4, no. 5, pp. 850–5, Oct. 1992.

[9] S. G. Thomas and V. E. Franklin-Tong, "Self-incompatibility triggers programmed cell death in Papaver pollen," *Nature*, vol. 429, no. 6989, pp. 305–309, May 2004.

[10] W. C. Fuqua, S. C. Winans, and E. P. Greenberg, "Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.," *J. Bacteriol.*, vol. 176, no. 2, pp. 269–75, Jan. 1994.

[11] P. Mehta, S. Goyal, T. Long, B. L. Bassler, and N. S. Wingreen, "Information processing and signal integration in bacterial quorum sensing," *Mol Syst Biol*, vol. 5, Nov. 2009.

[12] D. E. Cameron, C. J. Bashor, and J. J. Collins, "A brief history of synthetic biology," *Nat. Rev. Microbiol.*, vol. 12, no. 5, pp. 381–390, Apr. 2014.

[13] N. Anesiadis, W. R. Cluett, and R. Mahadevan, "Dynamic metabolic engineering for increasing bioprocess productivity," *Metab. Eng.*, vol. 10, no. 5, pp. 255–266, Sep. 2008.

[14] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging discipline.," *Mol. Syst. Biol.*, vol. 2, p. 2006.0028, 2006.

[15] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, “Synthetic biology: new engineering rules for an emerging discipline,” *Mol. Syst. Biol.*, vol. 2, May 2006.

[16] S. Elmore, “Apoptosis: a review of programmed cell death.,” *Toxicol. Pathol.*, vol. 35, no. 4, pp. 495–516, Jun. 2007.

[17] S. L. Harris and A. J. Levine, “The p53 pathway: positive and negative feedback loops,” *Oncogene*, vol. 24, no. 17, pp. 2899–2908, Apr. 2005.

[18] A. Brill, A. Torchinsky, H. Carp, and V. Toder, “The Role of Apoptosis in Normal and Abnormal Embryonic Development,” *J. Assist. Reprod. Genet.*, vol. 16, no. 10, pp. 512–519, 1999.

[19] C. Janeway, *Immunobiology 5 : the immune system in health and disease*. Garland Pub, 2001.

[20] G. P. Gupta, J. Massagué, J. M. Chirgwin, M. Dallas, B. G. Grubbs, R. Wieser, J. Massagué, G. R. Mundy, T. A. Guise, Y. Chen, and et al., “Cancer metastasis: building a framework.,” *Cell*, vol. 127, no. 4, pp. 679–95, Nov. 2006.

[21] A. F. Chambers, A. C. Groom, and I. C. MacDonald, “Metastasis: Dissemination and growth of cancer cells in metastatic sites,” *Nat. Rev. Cancer*, vol. 2, no. 8, pp. 563–572, Aug. 2002.

[22] A. K. Patnaik, W. J. Ehler, and E. G. MacEwen, “Canine Cutaneous Mast Cell Tumor: Morphologic Grading and Survival Time in 83 Dogs,” *Vet. Pathol.*, vol. 21, no. 5, pp. 469–474, Sep. 1984.

[23] C. A. London and B. Seguin, “Mast cell tumors in the dog.,” *Vet. Clin. North Am. Small Anim. Pract.*, vol. 33, no. 3, p. 473–89, v, May 2003.

[24] R. Kalluri, M. S. Adams, K. Fishwick, M. Bronner-Fraser, and M. A. Nieto, “EMT: when epithelial cells decide to become mesenchymal-like cells.,” *J. Clin. Invest.*, vol. 119, no. 6, pp. 1417–9, Jun. 2009.

[25] B. M. Gumbiner, “Epithelial morphogenesis.,” *Cell*, vol. 69, no. 3, pp. 385–7, May 1992.

[26] R. Kalluri and R. A. Weinberg, “The basics of epithelial-mesenchymal transition.,” *J. Clin. Invest.*, vol. 119, no. 6, pp. 1420–8, Jun. 2009.

[27] I. Pastar, O. Stojadinovic, N. C. Yin, H. Ramirez, A. G. Nusbaum, A. Sawaya, S. B. Patel, L. Khalid, R. R. Isseroff, and M. Tomic-Canic, “Epithelialization in Wound Healing: A Comprehensive Review.,” *Adv. wound care*, vol. 3, no. 7, pp. 445–464, Jul. 2014.

[28] R. Kalluri and E. G. Neilson, “Epithelial-mesenchymal transition and its implications for fibrosis,” *J. Clin. Invest.*, vol. 112, no. 12, pp. 1776–1784, Dec. 2003.

[29] A. K. Kiemer, K. Takeuchi, and M. P. Quinlan, “Identification of genes involved in epithelial-mesenchymal transition and tumor progression,” *Oncogene*, vol. 20, no. 46, pp. 6679–6688, Oct. 2001.

[30] J. P. Thiery, “Epithelial–mesenchymal transitions in tumour progression,” *Nat. Rev.*

Cancer, vol. 2, no. 6, pp. 442–454, Jun. 2002.

[31] E. W. Thompson, D. F. Newgreen, and D. Tarin, “Carcinoma Invasion and Metastasis: A Role for Epithelial-Mesenchymal Transition?,” *Cancer Res.*, vol. 65, no. 14, pp. 5991–5995, Jul. 2005.

[32] M. G. Mendez, S.-I. Kojima, and R. D. Goldman, “Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition.,” *FASEB J.*, vol. 24, no. 6, pp. 1838–51, Jun. 2010.

[33] E. D. Hay, “The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it,” *Dev. Dyn.*, vol. 233, no. 3, pp. 706–720, Jul. 2005.

[34] H. Tsukamoto, K. Shibata, H. Kajiyama, M. Terauchi, A. Nawa, F. Kikkawa, and et al., “Irradiation-induced epithelial-mesenchymal transition (EMT) related to invasive potential in endometrial carcinoma cells.,” *Gynecol. Oncol.*, vol. 107, no. 3, pp. 500–4, Dec. 2007.

[35] C. M. Nelson, D. Khauv, M. J. Bissell, and D. C. Radisky, “Change in cell shape is required for matrix metalloproteinase-induced epithelial-mesenchymal transition of mammary epithelial cells.,” *J. Cell. Biochem.*, vol. 105, no. 1, pp. 25–33, Sep. 2008.

[36] S. E. Leggett, J. Y. Sim, J. E. Rubins, Z. J. Neronha, E. K. Williams, and I. Y. Wong, “Morphological single cell profiling of the epithelial-mesenchymal transition.,” *Integr. Biol. (Camb).*, vol. 8, no. 11, pp. 1133–1144, Nov. 2016.

[37] M. Yilmaz and G. Christofori, “EMT, the cytoskeleton, and cancer cell invasion,” *Cancer Metastasis Rev.*, vol. 28, no. 1–2, pp. 15–33, Jun. 2009.

[38] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, “Histopathological image analysis: a review.,” *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–71, 2009.

[39] Y. Hu, S. Xu, W. Jin, Q. Yi, and W. Wei, “Effect of the PTEN gene on adhesion, invasion and metastasis of osteosarcoma cells.,” *Oncol. Rep.*, vol. 32, no. 4, pp. 1741–7, Oct. 2014.

[40] Z. Yin, A. Sadok, H. Sailem, A. McCarthy, X. Xia, and F. Li, “A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes,” *Nat. cell*, 2013.

[41] C. M. Fillmore and C. Kuperwasser, “Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy.,” *Breast Cancer Res.*, vol. 10, no. 2, p. R25, 2008.

CHAPTER 2: MATHEMATICAL MODELING APPLICATIONS IN CELLULAR BIOLOGY

2.1 Modeling in biology

The partnership between modeling and experiment has always been central to the advancement of understanding in chemistry and physics, however the widespread adoption of mathematical models in biology has been slower to develop. This is due in part to the complexity and redundancy of biological systems. The recent advancement in biological models is multifactorial. Key developments which contributed to this advancement include: the dramatic increase in data from novel high throughput biological technologies including proteomics and genomics [1] advancing modeling techniques such as machine learning and the evolving field of data science [2], increasing multidisciplinary collaboration including funding focused on this type of research [3], and consistent demonstration of the utility of models [4][5]. The symbiotic relationship between computational modeling and biological experiment is a unifying theme of this thesis and the evolution of this relationship is expanded upon on in later chapters.

2.1.1 Application of biological modeling

The applications of biological models are diverse. Simple models can be thought provoking to intuit deeper understanding of biological processes, used to validate hypotheses, and can guide development of experimentally testable predictions. As models increase in complexity their role may expand, but at their core, the uses to validate and predict persist. More complex models may be used for sensitivity analysis [6], to understand cellular networks and processes as systems [7], to develop useful synthetic systems that serve a real-world function [8], [9], and to predict complex biological behavior in a clinically meaningful way [5].

2.1.2 Value of models

The value of modeling cannot be understated. When used cohesively with experiment, models have recently resulted in some incredible, front-page-worthy developments – such as reprogramming immune cells to kill cancer [10] or engineering a logic circuit to detect cancer cells and force them to undergo apoptosis [11]. Models also are integrative to the development of important pharmaceuticals and other biologic compounds such as *in silico* driven drug design [12], development of synthetic biology systems to produce pharmaceuticals such as antimalarial drugs and chemotherapeutic agent taxol [13], [14], and optimization of biofuel production [15].

There are both intrinsic and extrinsic values to modeling. For example, a model can be used to refine an experimental design to a small subset of variables to test thus saving valuable resources. This technique was key to the success of the Medford lab's development of a synthetic signaling pathway in plants [9]. By using modeling to both characterize components and predict behavior when these components were assembled, the lab narrowed down a nearly infinite list of possible components and combinations into a specific engineered gene circuit, which saved unimaginable amounts of time and resources. In a research world where funding and time are ever in short supply, the utilization of modeling to refine an experiment to focus upon the precise conditions most likely to yield a successful result is invaluable. Time and money are not the only important resources modeling and improve upon. For experiments requiring animal models or for clinical trials for both animals and humans, there is an added ethical benefit when modeling is able to reduce or replace the number of *in vivo* experiments needed to achieve the end goal [16], [17].

More intrinsic value can be seen through identification of gaps in knowledge and through gaining fundamental understanding of biological phenomena. The identification of gaps in our current understanding interestingly is often discovered when a model fails to match observed biologic behavior. There are many examples of scientists developing a model only to discover that the model does not replicate the measured biologic outcome. This leads to extensive thought and re-design of theoretical models to explain this mismatch. In some cases, this iterative modeling resulted in a prediction for a previously unknown genetic network interaction which was later found to exist experimentally [18].

When properly designed and applied, models can serve a myriad of roles from giving researchers detailed instructions on where to look and what to look for, a “reality check” of our assumptions [1], decreasing resources through the use of faster and less expensive *in silico* experiments, and allowing for the development of complex systems such as those used in synthetic biology.

2.2 Building and development of models

The question of how to develop a biological model is a complex one and there are as many approaches as there are techniques. Once a problem, system or question to model has been selected, the framework for the theoretical model must be chosen along with the scope of the model. This is inherently tied to the goal of the model and the questions one wishes to ask. In building the model, rarely are all interactions known a priori and so one must theorize connections from other experiments and intuitive understanding of the process being modeled. Parameters are selected based upon measured literature values when known, and more often, best estimates based on similar known values or through computational optimization algorithms designed to screen parameters and model outputs with known experimental outputs. Once a model is constructed, it is typically compared with experiment in an iterative process, focusing on where the model and experiment align or diverge. This allows us to update our understanding and evolve a better model.

Here, we discuss a few strategies commonly employed, each carries challenges and benefits. Design of any type of model may be loosely broken down into two strategies, based upon the point of view. A bottom-up approach works by building small, detailed systems and combining them into more complex systems; this technique typically requires the base systems, or small building blocks, be characterized in great detail. Conversely, a top-down approach starts by formulating an overview, but does not necessarily incorporate details of subsystem levels, and lends itself more to theoretically based models where details are filled in over time.

In a bottom-up model, building blocks are typically molecules, proteins or DNA/RNA. Interactions between these components can be modeled with biochemical kinetics, such as mass action or Michaelis-Menten rate laws. In the case of large concentrations of components, the model can be specified with ordinary differential equations (ODEs). The set of equations forms a dynamical system that can be analyzed qualitatively for its global properties or simulated. When molecule numbers are small, the appropriate choice is a stochastic representation of the biochemical processes utilizing a chemical master equation. Most often the chemical master equation cannot be solved analytically so either numerical approximations or stochastic simulations such as Monte Carlo or Gillespie iterative simulations are used.

With a top-down overview, small details of individual molecules and proteins are often not specified. Rather, during development of a top-down model, a 'black box' depiction of individual mechanisms is commonly used, with the assumption that the simplification does not significantly alter the accuracy of the model. This application can be used for phenomenological models where the underlying mechanism is not well understood, such as a theoretical model of the network control of the differentiation of bone marrow stromal cells into bone and muscle, discussed below [41]. Another use is with systems data based models, where the explanation may lie somewhere in the data set but current understanding lacks a mechanistic explanation; we can then use machine learning or similar computational approaches to build predictive models.

2.2.1 Bottom-up models

Bottom-up systems biology many times begins with known molecular interactions, requiring characterization of single molecular properties and interactions. A common objective is to integrate the smaller pathway models into larger scale models which can simulate the entire system. Some examples of this approach include characterization of the signaling network of the epidermal growth factor receptor (EGF) [19]–[21] and the TGF-beta signaling pathway [22], [23]. A different application of this technique can be found in drug design optimization, with far-reaching implications such as personalized and predictive medicine, based upon detailed understanding of molecular-level interactions and use of individual molecules as biomarkers [24], [25]. One example of this, utilizing this well-characterized EGF pathway, is identification of specific anti-cancer therapies as well as biomarkers to guide predictions in the likelihood of success of those therapies [26], [27].

The pathway, which was initially understood through a bottom-up, detailed approach [19], [20], [28], [29], was proposed as a target for cancer therapy due to the observation that the receptor was highly expressed in epithelial cancers. After extensive research, which is well-summarized in ref: [30], [31] the understanding of the EGFR pathway mechanisms has grown further. Advances include mechanistic understanding of acquired resistance to chemotherapeutic agents [32]–[34], and biomarkers indicative of higher response to chemotherapeutic agents (anti-EGFR drugs gefitinib and erlotinib) [26], [27]. This intensive, collaborative research, which was rooted in decades long accumulation of understanding of small signaling interactions and protein structure, grew from a bottom-up direction to result in two classes of actively used cancer treatments: specific antibodies targeting the extracellular domain of the EGF receptor and competitive inhibitors of the tyrosine kinase of the receptor. These treatments are being used in colorectal, neck, head, lung and pancreatic cancers [35], [36].

One example of the successful combination of modeling and bottom up detailed understanding of the EGFR pathway is Ref. [37]. The authors built upon the signaling network characterized by many and modeled by Kholodenko et al [38], which analyzed three coupled cycles of protein interactions of specific phosphotyrosine residues on the EGF receptor in 23 coupled ordinary differential equations. Araujo et al. built upon this model and mathematically simulated the effect of small molecule tyrosine kinase inhibitors targeted at specific 'network nodes', meaning inhibition of a given interaction such as phosphorylation, de-phosphorylation, binding or dissociation. They simulated the effect of these inhibitors on output of the most downstream variable, three activated cytoplasmic proteins: Grb2, Shc which act to initiate the membrane-bound Ras protein and advance cell cycle; and PLC-gamma which can stimulate cell motility [39]. Inhibition was simulated by decreasing the forward rate constant from 0 to 90%. The authors discovered that downstream signals of inhibition are enhanced when multiple upstream processes are inhibited, allowing for lower doses of potentially systemically toxic medications. Notably, this effect is seen the most when the targets are serially connected, that is subsequent interactions rather than parallel or unrelated interactions. The most clinically relevant conclusion from this work, is that combination drug therapy targeting multiple, consecutive nodes of a signaling cascade can have a supra-additive effect, allowing for much lower doses of any given drug. This is important as many anti-cancer drugs, even those specifically targeted to one signaling pathway, have profound negative side effects.

2.2.2 Top-down models

The mitogen-activated protein kinase (MAPK) pathway is necessary for differentiation of precursor cells into muscle [40]. Wang et all observed that for the class of mesenchymal progenitor cell studied, when cultured with an inhibitor of MAPK and BMP2, a protein needed to induce osteogenic (bone) differentiation, the cells could differentiate into bone [41]. Moreover, the progenitor cells demonstrated osteogenic differentiation in a switch-like manner, meaning that at a concentration of BMP2 <150 ng/mL, none of the cells demonstrated markers of bone, and for concentrations > 200 ng/mL, nearly all cells demonstrated markers of bone. This non-linear, 'ultrasensitive' response is characteristic of a bistable switch. This observation let the authors to develop a mathematical model of differential equations, based on a positive feedback loop. The model was then applied to analyze the differentiation of cells in response to the BMP2 concentrations and found their experimental data had extremely close agreement with a stochastic simulation of the model. Lastly, the authors successfully used the model to make a prediction that was tested experimentally and found to align with the model. Their model predicted that past stimulation could activate the positive feedback loop, which would alter a cell's response to future stimulation, that is, exhibit memory. Experimentally, they pretreated cells to stimulate them to differentiate into bone with BMP and the MAPK inhibitor for 7 days, and then re-plated and challenged the cells with varying doses of BMP2. The authors found that, as predicted by their model, the pretreated cells demonstrated osteogenic differentiation at much lower doses of BMP2, having the effect of making the switch easier to initiate. This work is an excellent example of a top-down model, where the specific gene circuits and dynamics of a positive feedback loop, as proposed by the authors, are not known, but a phenomenological model successfully explained observed cellular experimental data and was utilized to make predictions regarding further experimental data, which were found to be accurate. Next steps of this work include searching for the 'black box' of specific cellular protein and gene interactions that result in this positive feedback loop. A similar approach was used to characterize the bistable lactose network in *E. coli* [42] and to identify and model a gene regulatory network with bistability for the cellular differentiation of hematopoietic stem cells [43].

2.3 Specific Modeling Techniques

The range of computational modeling methods applied to biology is as varied as the applications and systems that are modeled. Inherent in model selection and design is parameter selection; strategies for identifying and estimating parameters are briefly discussed. Simple models can be constructed using ordinary differential equations to characterize genetic circuits or Monte Carlo simulations of kinetic parameters. More advanced modeling techniques can range from advanced statistics, complex metabolic modeling, and the use of machine learning for prediction and classification. Some successful applications of machine learning including to classify cancer cell types by gene expression and classify cell populations by shape are discussed.

2.3.1 Parameter search

A crucial component to an accurate mathematical model is selection of appropriate parameters. This selection is often a combination of utilizing known values directly measured by detailed experiments, estimated values based upon characterized reactions of similar signaling pathways, and optimization algorithms to identify a range for rate constants which result in experimentally observed biologic behavior. Much research has been done to develop methods to optimize parameter identification [6], [44], [45]. The TGF-beta signaling pathway, a large pathway with multiple branches and subgroups, is responsible for many important biological processes including differentiation during development and growth [46]. This pathway has been modeled with a set of deterministic ordinary differential equations successfully to deepen understanding of the pathway and the role of the various observed feedbacks in regulating this complex signaling pathway [47]–[52]. In addition to the equations themselves, the success of these deterministic models also depends upon selection of both the initial conditions and the rate constants for individual reactions. Some parameters, such as the cell size which determines reaction volume, can be exactly measured via confocal microscopy or can be approximated from calculations of cell diameter. Similarly, initial conditions can be determined through measurement of protein concentrations utilizing techniques such as immunoassay. Rate constants for some of the reactions have been measured experimentally [47], [52]–[54]. Others are determined through modeling of the system and applying optimization algorithms; this can be time consuming and computationally require running many simulations however there is an increasing number of parameter estimation tools available to expedite the process [55]–[58]. The TGF-beta pathway is a great example of the combination of these techniques which has led to detailed knowledge about model parameters for the canonical TGF-beta/Smad signaling pathway and is well reviewed [22].

One approach to modeling biology seeks to use simple mathematical models to represent the system. Ordinary differential equations describing the dynamics of transcriptional regulation have successfully been used in many examples. Often, basic kinetic rate-law and Michaelis-Menten equations are combined into a system of ordinary differential equations which capture the dynamics of the network. In many cases, these differential equations are simplified into dimensionless models. Many successful examples of this type of modeling, paired with experiments which correspond with model predictions, exist including: the Gardner synthetic genetic toggle switch of paired repressible promoters simulated with a derived mathematical model of dimensionless ordinary differential equations [59]; mathematical modeling of the mitogen-activated protein kinase (MAPK) cascade's ultrasensitive response ("all or none") with deterministic modeling of the positive feedback loop in *Xenopus* oocytes [60]; and a detailed mathematical model of the budding yeast cell cycle with nonlinear ordinary differential equations which accurately captures the time courses of three major classes of kinases involved in cell cycle control [61].

2.3.1.1 The Synthetic Repressilator: applications of deterministic and stochastic simple mathematical modeling

Yet another example of this is the synthetic repressilator designed by Elowitz and Leibler [62]. The authors developed a mathematical model of a series of three repressors connected in a negative feedback loop, each repressing the subsequent gene. They described this system with six coupled first-order differential equations which described the transcription, translation and degradation reactions of the repressor proteins (p) and their corresponding mRNA (m):

$$\frac{dm}{dt} = -m + \frac{\alpha}{(1 + p^n)} + \alpha_0$$

$$\frac{dp}{dt} = -\beta(p - m)$$

Where α_0 corresponds to leaky promoter transcription, $\alpha+ \alpha_0$ corresponds to un-repressed promoter transcription, β is the decay rate of the protein, and n is a Hill coefficient. The differential equations were simulated with a range of parameters, leading the authors to identify specific parameter combinations which would lead to specific steady state dynamics including and unstable steady state resulting in periodic oscillations of protein concentrations. These parameters indicated that oscillations occurred when strong promoters were utilized and there was a similar rate of mRNA and protein degradation. The authors then built the genetic circuit and showed that oscillations occurred only when their predicted parameter values were met, reading out the level of green fluorescent protein in individual cells which oscillated on and off as predicted.

Furthermore, the authors explored their model utilizing stochastic simulations of the repressilator to assess the effect of noise on the system, which revealed the amplitude of oscillations was much more variable when simulated stochastically, which was what was also observed experimentally, indicating that for their system size, noise and stochastic dynamics play a role in output, but also that the system can be qualitatively modeled deterministically.

2.3.1.2 Protein dynamic based mathematical modeling of cellular shape

Another example of a simple mathematical model utilizes protein dynamics rather than transcription dynamics. In Ref [63] the authors develop a mathematical model of Rac-Rho dynamics and its influence on cell shape. The Rho proteins constitute a family of small G proteins that play a major role in cytoskeletal dynamics and other important intracellular processes. Rho proteins are activated by guanine nucleotide exchange factors (GEFs) and are deactivated by the GTPase-activating proteins (GAPs) [64]. The prototypical Rho proteins, i.e. RhoA, RhoB and RhoC, share effectors and play similar roles. These Rho proteins activate Rho kinase or ROCK, which directly phosphorylates the myosin light chain, leading to activation of myosin. ROCK also inactivates myosin phosphatase, increasing myosin activation and thus actomyosin contractility. ROCK also phosphorylates LIM-kinase, which gets activated and then phosphorylates and inactivates cofilin, which is the protein that severs actin. Thus, Rho activation promotes actin polymerization and actomyosin contractility.

The Rac proteins are also a sub-family of the Rho proteins and include the proteins Rac1, Rac2, Rac3 and RhoG. Activated Rac has been found at the leading edge of migrating cells [65] and can activate actin assembly and lamellipodia formation through its activation of the WAVE complex that regulates actin polymerization and crosslinking [66].

The Holmes paper [63] draws on the experimental research that suggests that cells with high levels of Rac and little Rho are typically flat and spread out, while those that have high levels of Rho relative to Rac are rounded and contracted. Their literature survey suggests that Rho and Rac are often segregated in cells, with Rac being present at the lamellipodia and the leading edge, and Rho being present at the contractile training end of the cell. Note that this paper does not specify which Rho or Rac proteins are being considered prototypical, but the literature cited is mostly about RhoA and Rac1. The Holmes paper constructs two simplified phenomenological models about Rho-Rac interactions to see whether this basic model can reproduce the experimental observations. The basic model, which they call Model 2, includes activation steps for both Rho and Rac, and an inhibition step whereby active Rho inhibits the activation of Rac and vice versa. They also analyze another even simpler model called Model 1, where there is a positive feedback between the active GTPase and the activation-reaction of the inactive GTPase. Both these circuits have been studied previously, but one of the novel aspects of this paper is the spatial modeling of both circuits. It should be noted that Model 2 is somewhat similar to the toggle switch [59] in that it has two mutually repressing processes. However, no gene transcription is involved; all the processes are activation or deactivation signaling processes. Consequently, since GTPase activity usually takes place in seconds or at most minutes, it can be assumed that total GTPase is conserved. They also assume, based on previous work, that the diffusion constant of the inactive protein is much slower than that of the active protein. Both these assumptions are crucial for the results of their model. Their main aim is then to determine parameters that allow for the three outcomes: spreading (uniformly high Rac and low Rho); contraction (uniformly high Rho and low Rac) and polarization (spatially excluded high Rho and high Rac activities).

The most interesting result of their analysis is that they find that bistability can coexist with a stably static polarization pattern. This is surprising because it was previously assumed that bistability would lead to a uniformly high or low level of protein throughout the cell. However, their spatial model shows that in fact it is possible for bistability to coexist with polarization, with one protein being activated in the “front” of the cell, and another at the “rear”. Mechanistically this takes place because of the difference in diffusion constants, which can flip the switch towards the Rac side or the Rho side in one part of the cell, while maintaining it in the opposite state in the other part of the cell.

The simplicity of the Holmes model is a disadvantage when dealing with the diversity of cellular behavior but allows a more complete analysis of the model. This, as they have shown, can lead to some results that were not previously appreciated and underlines the power of phenomenological models as we have discussed elsewhere in this review.

2.3.2 Machine Learning

In addition to the traditional mathematical model applications already discussed, recently biologists have been incorporating increasingly advanced mathematical algorithms to gain insight into biological processes, to classify data, and to guide development of models and make predictions. One example is the use of machine learning - the use of algorithms to classify or recognize patterns and make predictions based on models built from experimental data [2]. This process involves establishing a model or classifier by computer/mechanized optimization rather than human design, and apply this computing machine to classify or characterize data. Some examples of its use are to identify translation initiation sites in *E. coli* [67], classify types of cancer [68], [69], predict protein structure [70], and even predict the development of PTSD based on early trauma [71]. Two detailed examples of the use of this technology are discussed: classification of cells based upon shape [72] and prediction of clinical outcome of diffuse large B-cell lymphoma based upon microarray data [5].

Yin et al. utilized machine learning to classify cells into discrete shape groups allowing them to ask many interesting questions about cell shape distribution in wild-type cells and in melanoma, as well as the gene networks regulating these cell shapes [72]. The authors measured or calculated 211 cell morphology features of an individual cell including wavelet transformations, whole-cell geometric measures, Zernike moments, and regional geometric measurements extracted from divided parts of cell segments. These features were utilized to model phenotype and classify cells into distinct cell shapes. Of these features, the top 20 most informative were selected through a Support Vector Machine recursive feature elimination method and this selection was further refined with a genetic algorithm support vector machine. These machine learning algorithms were successful in segregating *Drosophila* hemocyte cells into five discrete shapes, as validated by histopathologist classification of the cells. The authors then utilized this classification scheme to probe the effect of an array of genes affecting cell shape through an RNAi screen of genes and subclassified these genes into seven different classes based upon their effect on cell shape, concluding that most genes regulate transition between shapes rather than generating new shapes. They also applied their algorithm to develop a model of the transition of cells between shapes and concluded that the wild-type *Drosophila* cells studied made discrete switch-like transitions between shapes rather than a continuous transition with numerous intermediate shapes. That model of switch-like transition was also applied to a line of human melanoma cells that were also found to make a rapid switch-like conversion between shapes. This led to further work investigating the role specific signaling pathways, including tumor suppressor PTEN, in this transition and in those pathways' roles in determining cell shape. Specifically, they observed that PTEN and related genes decrease the variability of the population, leading to a population with one primary population of rounded or elongated cells. The advanced methods utilized by this group combined experiment with development of machine learning models to gain insight into signaling pathways governing cell shape and the mechanism by which cells change shape.

Shipp et al. identified a void in clinical outcomes models for diffuse large B-cell lymphoma (DLBCL), the most common lymphoid cancer in adults, and utilized machine learning to develop an algorithm to supplement current clinical decision-making protocols [5]. Prior to their work, prognosis for adults diagnosed with DLBCL was based upon age, response to treatment, stage, and a serum protein, through the International Prognostic Index (IPI) which was used to screen for patients unlikely to achieve cure with the standard therapy [73]. Utilizing RNA and oligonucleotide microarray data, the group developed a supervised machine learning algorithm to successfully differentiate DLBCL from the closely related B-cell lymphoma, follicular lymphoma (FL) [74]. This algorithm was accurate in 91% of the samples. A similarly structured algorithm was developed to predict the clinical outcome of patients with DLBCL. Data on the long-term clinical outcomes of the DLBCL patients in the prior study [74] was used to classify each case based on those with cured disease vs those with refractory or fatal disease

[5]. A supervised learning classification approach was again used to develop an outcome predictor. To assess the accuracy of the model the authors compared the results of the predictor with Kaplan-Meier survival analyses and found that the patients predicted to be cured by the machine had significantly longer survival. They further compared their findings to a prior study by Alizadeh et al [68], and although the comparison was limited as there was minimal overlap in microarray data between the two datasets and tumor samples, they were able to identify two genetic isoforms correlated with outcome from an independent data set. This work is just one example of many groups working to pair microarray screens of cancer with advanced computational algorithms to classify and make predictions for clinical decision making including optimization of cancer chemotherapy selection applying multiple types of machine learning algorithms [75], classification of lung cancer type based upon gene expression profile [76], and classification of skin cancer type through automated image analysis [77].

2.4 Modeling of Cellular Decisions with switches

Computational systems biology has been inspired by the argument that cellular phenotypes are stable attractors of the dynamics of the genetic and signaling network of cells. This argument was associated with the name of Kaufman [78], [79], but its roots can be found in previous work of Jacob and Monod as well [80]. An older but very powerful analogy was provided by Waddington, who coined the metaphor of cellular lineage specification during development as a particle finding a lowest energy configuration in a landscape of hills and valleys, representing cell fates and barriers between lineages [81].

As the complex transcriptional dynamics underlying cellular lineage specification became uncovered, a mechanistic basis for these landscapes became apparent. However actually linking specific networks of genes to attractors in some state space has proved more difficult. The most popular subnetworks understood have typically been switches based on either positive feedback or mutual inhibition. Because of the ubiquity of feedback, several putative switch-like topologies have been identified in the transcriptional and signaling network. Thus, lineage commitment has been often regarded as proceeding through a number of digital decisions, such as those between osteogenesis and chondrogenesis for mesenchymal stem cells [41].

The differentiation of blood cells into either the erythroid/megakaryocyte or the myelomonocytic lineage has been a much-studied model for lineage specification and it has given rise to some intriguing and influential mathematical modeling [82]. Two transcription factors that play key roles in this process are GATA1, a zinc finger TF, and PU1, a TF belonging to the Ets family. GATA1 is expressed in the erythroid lineage and PU1 in the myelomonocytic lineage. Moreover, both TF's repress each other and show a positive feedback with respect to their own transcription. In the absence of mathematical modeling, such a genetic architecture may lead to the simple assumption that the role of the positive feedback is to reinforce the lineage commitment to one lineage or the other. However, in Ref. [82] the authors construct and analyze a simple mathematical model of the basic decision circuit described briefly above, based on previous work in Ref. [83]. A cartoon of the model is shown in Figure 2.1 below.

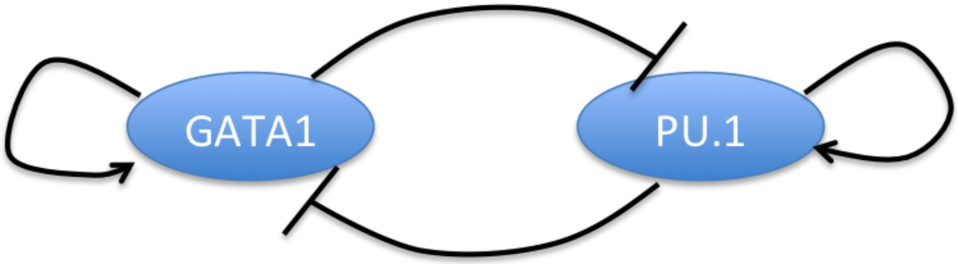


Figure 2.1 Representation of the multi-stable switch of GATA1 and PU.1

The mathematical model built was a phenomenological model in the sense that instead of trying to model mechanistic processes at the molecular level, the model assumed that a number of complex processes could be lumped into simpler phenomenological equations. In particular it modeled both the stimulatory and the inhibitory influences as sigmoidal relationships using Hill functions, which is standard for transcriptional processes, and well as processes involving multi-site phosphorylation [84]. The basic equations describing the relationships in Figure 2.1 are then:

$$\frac{dx_1}{dt} = a_1 \frac{x_1^n}{\theta_{a_1}^n + x_1^n} + b_1 \frac{\theta_{b_1}^n}{\theta_{b_1}^n + x_2^n} - k_1 x_1$$

$$\frac{dx_2}{dt} = a_2 \frac{x_2^n}{\theta_{a_2}^n + x_2^n} + b_2 \frac{\theta_{b_2}^n}{\theta_{b_2}^n + x_1^n} - k_2 x_2$$

Here the level of GATA1 is represented by the variable x1 and PU.1 is x2. The two activating Hill functions in each differential equation represent the positive feedback of x1 or x2 on itself, while the repressing Hill function represents the inhibition of x2 on x1 and x1 on x2 respectively. When autoregulation is not considered, this system reduces to the well-known genetic toggle switch. The genetic toggle switch has, with the right parameter values, two stable states, one corresponding to high x1 and low x2, and the other corresponding with low x1 and high x2. However, when the positive feedback is turned on we see a novel phenomenon – the emergence of three stable states! The third state that emerges has intermediate values of both x1 and x2 and lies somewhere on the 45-degree line when parameter values are symmetric. The authors argue that while the first two stable states that correspond in the biological system to high GATA1 and low PU.1 and low GATA1 and high PU.1 represent lineage commitment to either the erythroid or the myeloid state, the intermediate stable state represents a bipotential progenitor cell [82]. Parameter analysis of the system suggested that the parameter regime for the existence of tristable dynamics is quite broad, and thus tristability should be seen as the typical behavior of the circuit shown in Figure 2.1.

Ref. [82] leads to the intriguing question of how ubiquitous this tristable system is in stem cell differentiation or other processes. Named “the self-activating toggle switch (SATS)”, a number of significant papers have explored its role in the Epithelial-to-Mesenchymal transition in cancer [85]– [87]. This small but growing body of work has pointed out that the core regulatory circuit governing EMT has a similar attractor structure as the erythroid/myeloid transition, but is more complicated in appearance. Here the key regulatory unit is composed of two transcription factors, ZEB and SNAIL and two micro-RNA’s, miR-34 and miR-200 [86]. High levels of the two micro-RNAs are associated with the epithelial phenotype, while high levels of the SNAIL and ZEB are associated with the mesenchymal phenotype. The micro-RNAs repress both the transcription factors and are repressed by them. Moreover, these two transcriptional factors also show autoregulation, with Snail showing negative autoregulation and ZEB showing positive autoregulation. Despite the differences in the structure of the circuit however, mathematical analysis of this system showed that it was tristable, with the three stable states corresponding to the Epithelial phenotype, the Mesenchymal phenotype and a hybrid phenotype that had both epithelial and mesenchymal features. The existence of the hybrid explained some puzzling features in the data, and opened the door to new discoveries [87].

2.5 Application Based Modeling

Biological modeling is an iterative process which relies upon a close pairing with experiment for model refinement and identification of weaknesses in the model as well as testing of hypotheses generated by the model. These experiments can be *in vivo, in vitro* or *in silico*. For example, repeated *in silico* trials can test the importance of various parameters, leading to identification of targets for experimental work or deep understanding of gene regulator networks [6], [88]. Metabolic modeling can identify adjustments resulting in significant increases in yield of bioengineered products such as vitamins produced by bacteria, antibiotics produced by fungi, and biofuels produced by algae [89]. Synthetic biology has provided some great examples of this iterative process including bacteria which seeks out specific concentrations of molecules such as those given off by tumor cells [90], [91], genetic programs to induce apoptosis in cancer cells [11], and the engineering of plants to detect important molecules and alert to their presence [8], [9], [92].

2.5.1 Modeling in microbial cell factory design and optimization

An early synthetic biology application was the use of microbial cells as factories to produce products of importance. *S. cerevisiae* is one commonly used organism and the products produced by engineered *S. cerevisiae* are summarized in ref. [93]. Of note, human insulin [94] and several vaccines including those against human papilomavirus and hepatitis B virus are produced in *S. cerevisiae* [95], [96]. There are many microbes utilized for production of important products such as the production of hydrocodone from sugar in yeast [97], industrial enzymes and cell proteins in bacteria [98], or many different antibiotics in filamentous fungi (reviewed in ref. [99]). Other work has focused on the production of biofuels from microalgae, bacteria, filamentous fungi and yeasts through synthetic biology and engineered, modeled systems to optimize production systems that can be used in an industrial application [100], [101].

Many of these advances were made possible by biological modeling of the relevant systems. Insertion of a genetic pathway for creation of a product in a microbe does not automatically result in high yield of the desired product [102]. Much work has been done on methods to optimize cell factories including kinetic models, flux balance analysis and advanced metabolic modeling [103], [104].

One world-changing application of a harnessing the power of microbes to produce a product is the engineering of *E. coli* and *S. cerevisiae* to produce the gold standard treatment for malaria, which was initially cost prohibitive for many of those suffering from the disease and is now offered for free by many countries due to dramatically reduced costs. While mathematical modeling played a limited role in the direct design of the pathway to produce this medication, the foundation work characterizing metabolic networks of both host organisms was rooted in metabolic modeling including genome scale models and flux analyses.

Malaria is a deadly but treatable disease caused by *Plasmodium* parasites, transmitted by mosquitoes causing acute febrile illness which can lead to death. According to the World Health Organization, nearly half of the world's population was at risk of malaria with 91 countries found to have ongoing malaria transmission [105]. Artemisinin-based combination therapies (ACTs) are the gold standard treatment for malaria. Artemisinic acid is produced naturally by the *Artemisia annua* plant. While the chemical synthesis of this compound exists, it is too costly to produce on large scales and its extraction from the plant requires a large volume of plant material and large farming operations, which made the drug cost prohibitive for many [106]. This led to research into utilizing a host organism for the manufacturing of artemisinin through the highly collaborative, interdisciplinary Semi-synthetic Artemisinin Project [107].

The objective of the project was to engineer a microbe to produce an advanced precursor to artemisinin which is then converted into artemisinin via synthetic organic chemistry. Initially, *E. coli* was utilized to produce a precursor from acetyl-CoA [108]; further research identified a key enzyme in conversion was only found in eukaryotic cells and the host organism was switched to *Saccharomyces cerevisiae,* or baker's yeast [109]. The engineered yeast was capable of producing artemisinic acid, the immediate precursor to Artemisinin, by careful genetic engineering with synthetic regulation of some naturally occurring genes in the mevalonate pathway through upregulation and repression of specific genes, as well as introduction of a gene encoding the synthesis of a novel p450 enzyme from the plant *A. annua*. The engineered yeast is capable of producing artemisinic acid at a 500-fold increase over the plant [109].

While the development of the metabolic pathway to produce artemisinin in both *E. coli* and *S. cereviae* did not explicitly rely upon metabolic pathway modeling, the use of both microbes as a host organism for production was made possible by extensive characterization of both organisms over many years. The ability to engineer these organisms due to the extensive modeling that has already been done makes them an attractive and obvious choice. One tool utilized in metabolic engineering is genome-scale *in silico* metabolic models (GEM) which can predict engineering strategies and allow for rational metabolic engineering [89]. *E. coli* and *S. cerevisiae* are two of the most well studied and characterized microbial species. *E. coli*'s metabolic genome-scale model was first developed in the 90's and has been updated many times; it has been successfully applied to increase production of many compounds [89], [110], [111]. Additionally, the models have been used to identify lethal knockouts which expanded the number of knockout candidate mutants that could serve as a basis for design of new metabolite production by eliminating competing pathways. Similarly, the *S. cerevisiae* metabolic network was developed in 2003 [112] and updated through collaborative contributions. Numerous examples of successful application of genome-scale models for optimization of microbial production of pharmaceuticals, chemicals, enzymes, biofuels, food ingredients, and nutritional compounds and vitamins are reviewed in ref. [89].

2.6 Summary

As discussed, the importance of computational modeling has been and continues to be increasingly accepted in biology. Because of the massive influx of available data from advanced biological techniques, high-throughput “-omics”, and quantitative focused experimentation, development of high quality models has grown rapidly with endless examples of successful applications. Many biological models can be broken into one of two broad categories of models: a bottom-up model operating by piecing together details to give rise to a more complex model, starting with known interactions and parameters compared with a top-down model which provides an overview of the system without necessarily including details of first-level subsystems or interactions.

When used most effectively, a model is paired closely with experiment in biology to provide useful information. One example of this is the 'design-build-test-analyze' (DBTA) cycle utilized in synthetic biology applications such as biological engineering of microbes to produce useful products. Another example is the refinement of machine learning algorithms to improve accuracy in prediction of clinical outcome by testing the algorithm against increasingly larger datasets, which can identify successes or failures and result in model refinement.

REFERENCES

[1] A. Levchenko, “Computational cell biology in the post-genomic era.,” *Mol. Biol. Rep.*, vol. 28, no. 2, pp. 83–9, 2001.

[2] A. L. Tarca, V. J. Carey, X. Chen, R. Romero, and S. Drăghici, “Machine Learning and Its Applications to Biology,” *PLoS Comput. Biol.*, vol. 3, no. 6, p. e116, 2007.

[3] P. A. Sharp and R. Langer, “Promoting Convergence in Biomedical Science,” *Science (80-.).*, vol. 333, no. 6042, 2011.

[4] P. E. M. Purnick and R. Weiss, “The second wave of synthetic biology: from modules to systems,” *Nat. Rev. Mol. Cell Biol.*, vol. 10, no. 6, pp. 410–422, Jun. 2009.

[5] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. T. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. S. Pinkus, T. S. Ray, M. A. Koval, K. W. Last, A. Norton, T. A. Lister, J. Mesirov, D. S. Neuberg, E. S. Lander, J. C. Aster, and T. R. Golub, “Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning,” *Nat. Med.*, vol. 8, no. 1, pp. 68–74, Jan. 2002.

[6] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, “A methodology for performing global uncertainty and sensitivity analysis in systems biology,” *J. Theor. Biol.*, vol. 254, no. 1, pp. 178–196, 2008.

[7] P. K. Kreeger and D. A. Lauffenburger, “Cancer systems biology: a network modeling perspective,” *Carcinogenesis*, vol. 31, no. 1, pp. 2–8, Jan. 2010.

[8] M. S. Antunes, K. J. Morey, J. J. Smith, K. D. Albrecht, T. A. Bowen, J. K. Zdunek, J. F. Troupe, M. J. Cuneo, C. T. Webb, H. W. Hellinga, and J. I. Medford, “Programmable Ligand Detection System in Plants through a Synthetic Signal Transduction Pathway,” *PLoS One*, vol. 6, no. 1, p. e16292, Jan. 2011.

[9] K. A. Schaumberg, M. S. Antunes, T. K. Kassaw, W. Xu, C. S. Zalewski, J. I. Medford, and A. Prasad, “Quantitative characterization of genetic parts and circuits for plant synthetic biology,” *Nat. Methods*, vol. 13, no. 1, pp. 94–100, Nov. 2015.

[10] C. C. Kloss, M. Condomines, M. Cartellieri, M. Bachmann, and M. Sadelain, “Combinatorial antigen recognition with balanced signaling promotes selective tumor eradication by engineered T cells,” *Nat. Biotechnol.*, vol. 31, no. 1, pp. 71–75, Dec. 2012.

[11] Z. Xie, L. Wroblewska, L. Prochazka, R. Weiss, and Y. Benenson, “Multi-input RNAi-based logic circuit for identification of specific cancer cells.,” *Science*, vol. 333, no. 6047, pp. 1307–11, Sep. 2011.

[12] S. Ekins, J. Mestres, and B. Testa, “In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling.,” *Br. J. Pharmacol.*, vol. 152, no. 1, pp. 9–20, Sep. 2007.

[13] D.-K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong, and J. D. Keasling, "Production of the antimalarial drug precursor artemisinic acid in engineered yeast.," *Nature*, vol. 440, no. 7086, pp. 940–3, Apr. 2006.

[14] P. K. Ajikumar, W.-H. Xiao, K. E. J. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer, and G. Stephanopoulos, "Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli," *Science (80-.).*, vol. 330, no. 6000, pp. 70–74, Oct. 2010.

[15] D. R. Georgianna and S. P. Mayfield, "Exploiting diversity and synthetic biology for the production of algal biofuels," *Nature*, vol. 488, no. 7411, pp. 329–335, Aug. 2012.

[16] V. Baumans, "Use of animals in experimental research: an ethical dilemma?," *Gene Ther.*, vol. 11, pp. S64–S66, Oct. 2004.

[17] A. Natsch, R. Emter, and G. Ellis, "Filling the Concept with Data: Integrating Data from Different In Vitro and In Silico Assays on Skin Sensitizers to Explore the Battery Approach for Animal-Free Skin Sensitization Testing," *Toxicol. Sci.*, vol. 107, no. 1, pp. 106–121, Jan. 2009.

[18] J. E. Ferrell, J. R. Pomerening, S. Y. Kim, N. B. Trunnell, W. Xiong, C.-Y. F. Huang, and E. M. Machleder, "Simple, realistic models of complex biological processes: Positive feedback and bistability in a cell fate switch and a cell cycle oscillator," *FEBS Lett.*, vol. 583, no. 24, pp. 3999–4005, Dec. 2009.

[19] A. Kiyatkin, E. Aksamitiene, N. I. Markevich, N. M. Borisov, J. B. Hoek, and B. N. Kholodenko, "Scaffolding protein Grb2-associated binder 1 sustains epidermal growth factor-induced mitogenic and survival signaling by multiple positive feedback loops.," *J. Biol. Chem.*, vol. 281, no. 29, pp. 19925–38, Jul. 2006.

[20] A. Suenaga, A. B. Kiyatkin, M. Hatakeyama, N. Futatsugi, N. Okimoto, Y. Hirano, T. Narumi, A. Kawai, R. Susukita, T. Koishi, H. Furusawa, K. Yasuoka, N. Takada, Y. Ohno, M. Taiji, T. Ebisuzaki, J. B. Hoek, A. Konagaya, and B. N. Kholodenko, "Tyr-317 Phosphorylation Increases Shc Structural Rigidity and Reduces Coupling of Domain Motions Remote from the Phosphorylation Site as Revealed by Molecular Dynamics Simulations," *J. Biol. Chem.*, vol. 279, no. 6, pp. 4657–4662, Nov. 2003.

[21] H. S. Wiley, S. Y. Shvartsman, and D. A. Lauffenburger, "Computational modeling of the EGF-receptor system: a paradigm for systems biology.," *Trends Cell Biol.*, vol. 13, no. 1, pp. 43–50, Jan. 2003.

[22] Z. Zi, D. A. Chapnick, and X. Liu, "Dynamics of TGF-β/Smad signaling," *FEBS Lett.*, vol. 586, no. 14, pp. 1921–1928, Jul. 2012.

[23] J. Massagué, "TGF-β SIGNAL TRANSDUCTION," *Annu. Rev. Biochem.*, vol. 67, no. 1, pp. 753–791, Jun. 1998.

[24] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood, "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.," *Science*, vol. 292, no. 5518, pp. 929–34, May 2001.

[25] A. B. Kantor, W. Wang, H. Lin, H. Govindarajan, M. Anderle, A. Perrone, and C. Becker, “Biomarker discovery by comprehensive phenotyping for autoimmune diseases,” *Clin. Immunol.*, vol. 111, no. 2, pp. 186–195, May 2004.

[26] J. G. Paez, P. A. Jänne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson, and M. Meyerson, “EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy,” *Science (80-.).*, vol. 304, no. 5676, 2004.

[27] W. Pao, V. Miller, M. Zakowski, J. Doherty, K. Politi, I. Sarkaria, B. Singh, R. Heelan, V. Rusch, L. Fulton, E. Mardis, D. Kupfer, R. Wilson, M. Kris, and H. Varmus, “EGF receptor gene mutations are common in lung cancers from &quot;never smokers&quot; and are associated with sensitivity of tumors to gefitinib and erlotinib.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 36, pp. 13306–11, Sep. 2004.

[28] Y. Yarden and M. X. Sliwkowski, “Untangling the ErbB signalling network.,” *Nat. Rev. Mol. Cell Biol.*, vol. 2, no. 2, pp. 127–137, Feb. 2001.

[29] N. E. Hynes and H. A. Lane, “ERBB receptors and cancer: the complexity of targeted inhibitors,” *Nat. Rev. Cancer*, vol. 5, no. 5, pp. 341–354, May 2005.

[30] N. Minc, D. Burgess, and F. Chang, “Influence of Cell Geometry on Division-Plane Positioning,” *Cell*, vol. 144, no. 3, pp. 414–426, Feb. 2011.

[31] S. R. Hubbard, “EGF receptor inhibition: Attacks on multiple fronts,” *Cancer Cell*, vol. 7, no. 4, pp. 287–288, Apr. 2005.

[32] S. Kobayashi, H. Ji, Y. Yuza, M. Meyerson, K.-K. Wong, D. G. Tenen, and B. Halmos, “An Alternative Inhibitor Overcomes Resistance Caused by a Mutation of the Epidermal Growth Factor Receptor,” *Cancer Res.*, vol. 65, no. 16, 2005.

[33] W. Pao, V. A. Miller, K. A. Politi, G. J. Riely, R. Somwar, M. F. Zakowski, M. G. Kris, and H. Varmus, “Acquired Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib Is Associated with a Second Mutation in the EGFR Kinase Domain,” *PLoS Med.*, vol. 2, no. 3, p. e73, Feb. 2005.

[34] T. A. Carter, L. M. Wodicka, N. P. Shah, A. M. Velasco, M. A. Fabian, D. K. Treiber, Z. V Milanov, C. E. Atteridge, W. H. Biggs, P. T. Edeen, M. Floyd, J. M. Ford, R. M. Grotzfeld, S. Herrgard, D. E. Insko, S. A. Mehta, H. K. Patel, W. Pao, C. L. Sawyers, H. Varmus, P. P. Zarrinkar, and D. J. Lockhart, “Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 31, pp. 11011–6, Aug. 2005.

[35] J. Baselga and C. L. Arteaga, “Critical update and emerging trends in epidermal growth factor receptor targeting in cancer.,” *J. Clin. Oncol.*, vol. 23, no. 11, pp. 2445–59, Apr. 2005.

[36] M. Scaltriti and J. Baselga, “The Epidermal Growth Factor Receptor Pathway: A Model for Targeted Therapy,” *Clin. Cancer Res.*, vol. 12, no. 18, 2006.

[37] R. P. Araujo, E. F. Petricoin, and L. A. Liotta, “A mathematical model of combination therapy using the EGFR signaling network,” *Biosystems*, vol. 80, no. 1, pp. 57–69, 2005.

[38] B. N. Kholodenko, O. V Demin, G. Moehren, and J. B. Hoek, “Quantification of short term signaling by the epidermal growth factor receptor.,” *J. Biol. Chem.*, vol. 274, no. 42, pp. 30169–81, Oct. 1999.

[39] Y. Yarden and B.-Z. Shilo, “SnapShot: EGFR Signaling Pathway,” *Cell*, vol. 131, no. 5, p. 1018.e1-1018.e2, Nov. 2007.

[40] A. Keren, Y. Tamir, and E. Bengal, “The p38 MAPK signaling pathway: A major regulator of skeletal muscle development,” *Mol. Cell. Endocrinol.*, vol. 252, no. 1–2, pp. 224–230, Jun. 2006.

[41] L. Wang, B. L. Walker, S. Iannaccone, D. Bhatt, P. J. Kennedy, and W. T. Tse, “Bistable switches control memory and plasticity in cellular differentiation.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 16, pp. 6638–43, Apr. 2009.

[42] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden, “Multistability in the lactose utilization network of Escherichia coli,” *Nature*, vol. 427, no. 6976, pp. 737–740, Feb. 2004.

[43] P. Laslo, C. J. Spooner, A. Warmflash, D. W. Lancki, H.-J. Lee, R. Sciammas, B. N. Gantner, A. R. Dinner, and H. Singh, “Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates,” *Cell*, vol. 126, no. 4, pp. 755–766, Aug. 2006.

[44] C. G. Moles, P. Mendes, and J. R. Banga, “Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods,” *Genome Res.*, vol. 13, no. 11, pp. 2467–2474, Nov. 2003.

[45] A. Kremling, S. Fischer, K. Gadkar, F. J. Doyle, T. Sauter, E. Bullinger, F. Allgöwer, and E. D. Gilles, “A Benchmark for Methods in Reverse Engineering and Model Discrimination: Problem Formulation and Solutions,” *Genome Res.*, vol. 14, no. 9, pp. 1773–1785, Sep. 2004.

[46] Y. Shi and J. Massagué, “Mechanisms of TGF-beta signaling from cell membrane to the nucleus.,” *Cell*, vol. 113, no. 6, pp. 685–700, Jun. 2003.

[47] Z. Zi, Z. Feng, D. A. Chapnick, M. Dahl, D. Deng, E. Klipp, A. Moustakas, and X. Liu, “Quantitative analysis of transient and sustained transforming growth factor- signaling dynamics,” *Mol. Syst. Biol.*, vol. 7, no. 1, pp. 492–492, Apr. 2014.

[48] D. C. Clarke, M. D. Betterton, and X. Liu, “Systems theory of Smad signalling.,” *Syst. Biol. (Stevenage).*, vol. 153, no. 6, pp. 412–24, Nov. 2006.

[49] J. M. G. Vilar, R. Jansen, and C. Sander, “Signal Processing in the TGF-β Superfamily Ligand-Receptor Network,” *PLoS Comput. Biol.*, vol. 2, no. 1, p. e3, 2006.

[50] S. M. Lyons and A. Prasad, “Cross-talk and information transfer in mammalian and bacterial signaling.,” *PLoS One*, vol. 7, no. 4, p. e34488, 2012.

[51] P. Melke, H. Jönsson, E. Pardali, P. ten Dijke, and C. Peterson, “A Rate Equation Approach to Elucidate the Kinetics and Robustness of the TGF-β Pathway,” *Biophys. J.*, vol. 91, no. 12, pp. 4368–4380, Dec. 2006.

[52] B. Schmierer, A. L. Tournier, P. A. Bates, and C. S. Hill, “Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 18, pp. 6608–6613, May 2008.

[53] B. Schmierer and C. S. Hill, “Kinetic Analysis of Smad Nucleocytoplasmic Shuttling Reveals a Mechanism for Transforming Growth Factor -Dependent Nuclear Accumulation of Smads,” *Mol. Cell. Biol.*, vol. 25, no. 22, pp. 9845–9858, Oct. 2005.

[54] G. M. Di Guglielmo, C. Le Roy, A. F. Goodfellow, and J. L. Wrana, “Distinct endocytic pathways regulate TGF-β receptor signalling and turnover,” *Nat. Cell Biol.*, vol. 5, no. 5, pp. 410–421, May 2003.

[55] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, and SBML Forum, “The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.,” *Bioinformatics*, vol. 19, no. 4, pp. 524–31, Mar. 2003.

[56] Z. Zi and E. Klipp, “SBML-PET: a Systems Biology Markup Language-based parameter estimation tool,” *Bioinformatics*, vol. 22, no. 21, pp. 2704–2705, Nov. 2006.

[57] X. Ji and Y. Xu, “libSRES: a C library for stochastic ranking evolution strategy for parameter estimation,” *Bioinformatics*, vol. 22, no. 1, pp. 124–126, Jan. 2006.

[58] Z. Zi, “SBML-PET-MPI: a parallel parameter estimation tool for Systems Biology Markup Language based models,” *Bioinformatics*, vol. 27, no. 7, pp. 1028–1029, Apr. 2011.

[59] T. S. Gardner, C. R. Cantor, and J. J. Collins, “Construction of a genetic toggle switch in Escherichia coli.,” *Nature*, vol. 403, no. 6767, pp. 339–42, Jan. 2000.

[60] J. E. Ferrell and E. M. Machleder, “The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes.,” *Science*, vol. 280, no. 5365, pp. 895–8, May 1998.

[61] K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J. J. Tyson, “Kinetic analysis of a molecular model of the budding yeast cell cycle.,” *Mol. Biol. Cell*, vol. 11, no. 1, pp. 369–91, Jan. 2000.

[62] M. B. Elowitz and S. Leibler, “A synthetic oscillatory network of transcriptional regulators.,” *Nature*, vol. 403, no. 6767, pp. 335–8, Jan. 2000.

[63] W. R. Holmes and L. Edelstein-Keshet, “Analysis of a minimal Rho-GTPase circuit regulating cell shape,” *Phys. Biol.*, vol. 13, no. 4, p. 46001, Jul. 2016.

[64] S. Narumiya, M. Tanji, and T. Ishizaki, “Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion,” *Cancer Metastasis Rev.*, vol. 28, no. 1–2, pp. 65–76, Jun. 2009.

[65] E. E. Bosco, J. C. Mulloy, and Y. Zheng, “Rac1 GTPase: A ‘Rac’ of All Trades,” *Cell. Mol. Life Sci.*, vol. 66, no. 3, pp. 370–374, Feb. 2009.

[66] A. J. Ridley, “Rho GTPases and actin dynamics in membrane protrusions and vesicle trafficking,” *Trends Cell Biol.*, vol. 16, no. 10, pp. 522–529, Oct. 2006.

[67] G. D. Stormo, T. D. Schneider, L. Gold, and A. Ehrenfeucht, “Use of the ‘Perceptron’ algorithm to distinguish translational initiation sites in *E. coli*,” *Nucleic Acids Res.*, vol. 10, no. 9, pp. 2997–3011, 1982.

[68] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt, “Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling,” *Nature*, vol. 403, no. 6769, pp. 503–511, Feb. 2000.

[69] D. T. Ross, U. Scherf, M. B. Eisen, C. M. Perou, C. Rees, P. Spellman, V. Iyer, S. S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J. C. Lee, D. Lashkari, D. Shalon, T. G. Myers, J. N. Weinstein, D. Botstein, and P. O. Brown, “Systematic variation in gene expression patterns in human cancer cell lines.,” *Nat. Genet.*, vol. 24, no. 3, pp. 227–35, Mar. 2000.

[70] B. Rost and C. Sander, “Combining evolutionary information and neural networks to predict protein secondary structure,” *Proteins Struct. Funct. Genet.*, vol. 19, no. 1, pp. 55–72, May 1994.

[71] I. R. Galatzer-Levy, K.-I. Karstoft, A. Statnikov, and A. Y. Shalev, “Quantitative forecasting of PTSD from early trauma responses: A Machine Learning application,” *J. Psychiatr. Res.*, vol. 59, pp. 68–76, Dec. 2014.

[72] Z. Yin, A. Sadok, H. Sailem, A. McCarthy, X. Xia, and F. Li, “A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes,” *Nat. cell*, 2013.

[73] M. A. International Non-Hodgkin’s Lymphoma Prognostic Factors Project, D. P. Harrington, J. R. Anderson, J. O. Armitage, G. Bonadonna, G. Brittinger, F. Cabanillas, G. P. Canellos, B. Coiffier, J. M. Connors, R. A. Cowan, D. Crowther, S. Dahlberg, M. Engelhard, R. I. Fisher, C. Gisselbrecht, S. J. Horning, E. Lepage, T. A. Lister, J. H. Meerwaldt, E. Montserrat, N. I. Nissen, M. M. Oken, B. A. Peterson, C. Tondini, W. A. Velasquez, and B. Y. Yeap, “A predictive model for aggressive non-Hodgkin’s lymphoma.,” *N. Engl. J. Med.*, vol. 329, no. 14, pp. 987–94, 1993.

[74] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, “Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.,” *Science*, vol. 286, no. 5439, pp. 531–7, Oct. 1999.

[75] A. Petrovski, S. Shakya, and J. McCall, “Optimising cancer chemotherapy using an estimation of distribution algorithm and genetic algorithms,” in *Proceedings of the 8th annual conference on Genetic and evolutionary computation - GECCO ’06*, 2006, p. 413.

[76] M. D. Podolsky, A. A. Barchuk, V. I. Kuznetcov, N. F. Gusarova, V. S. Gaidukov, and S. A. Tarakanov, “Evaluation of Machine Learning Algorithm Utilization for Lung Cancer Classification Based on Gene Expression Levels.,” *Asian Pac. J. Cancer Prev.*, vol. 17, no. 2, pp. 835–8, 2016.

[77] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017.

[78] S. Huang, I. Ernberg, and S. Kauffman, “Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective,” *Semin. Cell Dev. Biol.*, vol. 20, no. 7, pp. 869–876, Sep. 2009.

[79] R. Zhu, A. S. Ribeiro, D. Salahub, and S. A. Kauffman, “Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models,” *J. Theor. Biol.*, vol. 246, no. 4, pp. 725–745, Jun. 2007.

[80] J. MONOD and F. JACOB, “Teleonomic mechanisms in cellular metabolism, growth, and differentiation.,” *Cold Spring Harb. Symp. Quant. Biol.*, vol. 26, pp. 389–401, 1961.

[81] J. E. Ferrell, “Bistability, Bifurcations, and Waddington’s Epigenetic Landscape,” *Curr. Biol.*, vol. 22, no. 11, pp. R458–R466, Jun. 2012.

[82] S. Huang, Y.-P. Guo, G. May, and T. Enver, “Bifurcation dynamics in lineage-commitment in bipotent progenitor cells,” *Dev. Biol.*, vol. 305, no. 2, pp. 695–713, May 2007.

[83] I. Roeder and I. Glauche, “Towards an understanding of lineage specification in hematopoietic stem cells: A mathematical model for the interaction of transcription factors GATA-1 and PU.1,” *J. Theor. Biol.*, vol. 241, no. 4, pp. 852–865, Aug. 2006.

[84] N. I. Markevich, J. B. Hoek, and B. N. Kholodenko, “Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades,” *J. Cell Biol.*, vol. 164, no. 3, pp. 353–359, Feb. 2004.

[85] D. Jia, M. K. Jolly, W. Harrison, M. Boareto, E. Ben-Jacob, and H. Levine, “Operating principles of tristable circuits regulating cellular differentiation,” *Phys. Biol.*, vol. 14, no. 3, p. 35007, May 2017.

[86] M. Lu, M. K. Jolly, H. Levine, J. N. Onuchic, and E. Ben-Jacob, “MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 45, pp. 18144–9, Nov. 2013.

[87] M. K. Jolly, S. C. Tripathi, J. A. Somarelli, S. M. Hanash, and H. Levine, “Epithelial-mesenchymal plasticity: How have quantitative mathematical models helped improve our understanding?,” *Mol. Oncol.*, May 2017.

[88] M. Oshiro, H. Shinto, Y. Tashiro, N. Miwa, T. Sekiguchi, M. Okamoto, A. Ishizaki, and K. Sonomoto, “Kinetic modeling and sensitivity analysis of xylose metabolism in Lactococcus lactis IO-1,” *J. Biosci. Bioeng.*, vol. 108, no. 5, pp. 376–384, 2009.

[89] C. B. Milne, P.-J. Kim, J. A. Eddy, and N. D. Price, “Accomplishments in genome-scale *in silico* modeling for industrial and medical biotechnology.,” *Biotechnol. J.*, vol. 4, no. 12, pp. 1653–70, Dec. 2009.

[90] S.-M. Tien, C.-Y. Hsu, B.-S. Chen, J. Armitage, J. Armitage, and N. Antonovsky, “Engineering Bacteria to Search for Specific Concentrations of Molecules by a Systematic Synthetic Biology Design Method,” *PLoS One*, vol. 11, no. 4, p. e0152146, Apr. 2016.

[91] J. C. Anderson, E. J. Clarke, A. P. Arkin, and C. A. Voigt, “Environmentally controlled invasion of cancer cells by engineered bacteria.,” *J. Mol. Biol.*, vol. 355, no. 4, pp. 619–27, Jan. 2006.

[92] M. S. Antunes, S.-B. Ha, N. Tewari-Singh, K. J. Morey, A. M. Trofka, P. Kugrens, M. Deyholos, and J. I. Medford, “A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity.,” *Plant Biotechnol. J.*, vol. 4, no. 6, pp. 605–22, Nov. 2006.

[93] I.-K. Kim, A. Roldão, V. Siewers, and J. Nielsen, “A systems-level approach for metabolic engineering of yeast cell factories,” *FEMS Yeast Res.*, vol. 12, no. 2, pp. 228–248, Mar. 2012.

[94] T. Kjeldsen, “Yeast secretory expression of insulin precursors.,” *Appl. Microbiol. Biotechnol.*, vol. 54, no. 3, pp. 277–86, Sep. 2000.

[95] E.-J. Kim, Y.-K. Park, H.-K. Lim, Y.-C. Park, and J.-H. Seo, “Expression of hepatitis B surface antigen S domain in recombinant Saccharomyces cerevisiae using GAL1 promoter,” *J. Biotechnol.*, vol. 141, no. 3–4, pp. 155–159, May 2009.

[96] H. J. Kim, S. J. Lee, and H.-J. Kim, “Optimizing the secondary structure of human papillomavirus type 16 L1 mRNA enhances L1 protein expression in Saccharomyces cerevisiae,” *J. Biotechnol.*, vol. 150, no. 1, pp. 31–36, Oct. 2010.

[97] S. Galanie, K. Thodey, I. J. Trenchard, M. Filsinger Interrante, and C. D. Smolke, “Complete biosynthesis of opioids in yeast,” *Science (80-.).*, vol. 349, no. 6252, pp. 1095–1100, Sep. 2015.

[98] J. C. Zweers, I. Barák, D. Becher, A. J. Driessen, M. Hecker, V. P. Kontinen, M. J. Saller, L. Vavrová, and J. M. van Dijl, “Towards the development of Bacillus subtilis as a cell factory for membrane proteins and protein complexes.,” *Microb. Cell Fact.*, vol. 7, p. 10, Apr. 2008.

[99] V. Meyer, “Genetic engineering of filamentous fungi — Progress, obstacles and future trends,” *Biotechnol. Adv.*, vol. 26, no. 2, pp. 177–185, Mar. 2008.

[100] V. L. Colin, A. Rodríguez, and H. A. Cristóbal, “The role of synthetic biology in the design of microbial cell factories for biofuel production.,” *J. Biomed. Biotechnol.*, vol. 2011, p. 601834, 2011.

[101] L. d’Espaux, D. Mendez-Perez, R. Li, and J. D. Keasling, “Synthetic biology for microbial production of lipid-based biofuels,” *Curr. Opin. Chem. Biol.*, vol. 29, pp. 58–65, 2015.

[102] J. Nielsen and M. C. Jewett, “Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*,” *FEMS Yeast Res.*, vol. 8, no. 1, pp. 122–131, Feb. 2008.

[103] J. Almquist, M. Cvijovic, V. Hatzimanikatis, J. Nielsen, and M. Jirstrand, “Kinetic models in industrial biotechnology – Improving cell factory performance,” *Metab. Eng.*, vol. 24, pp. 38–60, 2014.

[104] A. K. Fisher, B. G. Freedman, D. R. Bevan, and R. S. Senger, “A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories.,” *Comput. Struct. Biotechnol. J.*, vol. 11, no. 18, pp. 91–9, Aug. 2014.

[105] “World Malaria Report,” 2016.

[106] N. J. White, “Qinghaosu (Artemisinin): The Price of Success,” *Science (80-.).*, vol. 320, no. 5874, pp. 330–334, Apr. 2008.

[107] V. Hale, J. D. Keasling, N. Renninger, and T. T. Diagana, “Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs.,” *Am. J. Trop. Med. Hyg.*, vol. 77, no. 6 Suppl, pp. 198–202, Dec. 2007.

[108] V. J. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman, and J. D. Keasling, “Engineering a mevalonate pathway in Escherichia coli for production of terpenoids.,” *Nat. Biotechnol.*, vol. 21, no. 7, pp. 796–802, Jul. 2003.

[109] D.-K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong, and J. D. Keasling, “Production of the antimalarial drug precursor artemisinic acid in engineered yeast,” *Nature*, vol. 440, no. 7086, pp. 940–943, Apr. 2006.

[110] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson, “An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).,” *Genome Biol.*, vol. 4, no. 9, p. R54, 2003.

[111] J. L. Reed and B. Ø. Palsson, “Thirteen years of building constraint-based *in silico* models of Escherichia coli.,” *J. Bacteriol.*, vol. 185, no. 9, pp. 2692–9, May 2003.

[112] J. Förster, I. Famili, P. Fu, B. Ø. Palsson, and J. Nielsen, “Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network.,” *Genome Res.*, vol. 13, no. 2, pp. 244–53, Feb. 2003.

CHAPTER 3: CROSS-TALK AND INFORMATION TRANSFER IN MAMMALIAN AND BACTERIAL SIGNALING[1]

3.1 Summary

In mammalian and bacterial cells simple phosphorylation circuits play an important role in signaling. Bacteria have hundreds of two-component signaling systems that involve phosphotransfer between a receptor and a response regulator. In mammalian cells a similar pathway is the TGF-beta pathway, where extracellular TGF-beta ligands activate cell surface receptors that phosphorylate Smad proteins, which in turn activate many genes. In TGF-beta signaling the multiplicity of ligands begs the question as to whether cells can distinguish signals coming from different ligands, but transduced through a small set of Smads. Here we use information theory with stochastic simulations of networks to address this question. We find that when signals are transduced through only one Smad, the cell cannot distinguish between different levels of the external ligands. Increasing the number of Smads from one to two significantly improves information transmission as well as the ability to discriminate between ligands. Surprisingly, both total information transmitted and the capacity to discriminate between ligands are quite insensitive to high levels of cross-talk between the two Smads. Robustness against cross-talk requires that the average amplitude of the signals are large. We find that smaller systems, as exemplified by some two-component systems in bacteria, are significantly much less robust against cross-talk. For such system sizes phosphotransfer is also less robust

[1] This work has been published in PLoS one and is reproduced here under the Creative Commons License. Lyons, Samanthe M., and Ashok Prasad. "Cross-talk and information transfer in mammalian and bacterial signaling." *PloS one* 7.4 (2012): e34488.

against cross-talk than phosphorylation. This suggests that mammalian signal transduction can tolerate a high amount of cross-talk without degrading information content. This may have played a role in the evolution of new functionalities from small mutations in signaling pathways, allowed for the development of cross-regulation and led to increased overall robustness due to redundancy in signaling pathways. On the other hand the lack of cross-regulation observed in many bacterial two-component systems may partly be due to the loss of information content due to cross-talk.

3.2 Introduction

Phosphorylation reactions make up a large part of signal transduction processes. However there are many different topologies of phosphorylation-based signal transduction systems. In mammalian cells one of the simplest signal transduction networks is TGF-β signaling. TGF- β family members constitute a large class of related secreted polypeptides that are very important, especially during growth and development processes [1]. These proteins have been classified into several sub-families, of which the TGF- β subfamily of TGF- β's 1, 2 and 3 and the BMP sub-family, consisting of BMPs 2, 4, 5, 6, 8 and 9, is the most important. TGF- β family proteins signal through trans-membrane serine/threonine kinases known as Type I and Type II receptors. The TGF- β subfamily promotes the formation of a Type I/Type II complex after binding, while the BMP subfamily is believed to bind to a preformed complex of Type I/Type II receptors [2]. In either case, binding leads to phosphorylation of the cytoplasmic tail of the Type I receptor by the Type II receptor. The phosphorylated Type I receptor then recruits a subfamily of Smad proteins, called receptor Smads (or RSmads), that are phosphorylated by the Type I receptor. The 5 RSmads are the only known direct effectors of the TGF- β family of proteins and of them, Smad 1, 5 and 8 are preferentially used by BMP sub-family signaling and Smad 2 and 3 by the

TGF- β subfamily. Smad 4 is called a CoSmad and it binds with the phosphorylated RSmads and facilitates nuclear import. Smads 6 and 7 are a class of Smads called inhibitory Smads, or ISmads, and they negatively regulate Smad signaling [1].

Since the TGF- proteins are involved in diverse cellular and developmental processes, and the many proteins play non-redundant functions *in vivo*, the simple topology of the signaling pathway begs the question as to how specificity of signaling is maintained. The BMP subfamily, for example, can be divided further into two smaller families based on amino acid similarity, one containing BMP2 and BMP4 and the other the remaining BMPs. There is significant amino acid similarity within the BMP subfamily members, and even between subfamilies, but evidence suggests that they play non-redundant roles *in vivo* [3], [4], suggesting that the cell must be able to distinguish between the signals emanating from different BMP ligands.

Ligands in the extra-cellular space appear to preferentially bind different classes of receptors, and in particular it has been shown that BMPs 2/4 preferentially utilize the Type I receptor BMPR1A and the Type II receptor BMPR2 while BMPs 6/7 preferentially utilize ACVR1A and ACVR2A [5], but as far as is known they both signal through the same set of receptor Smads. It is therefore not clear whether the cell can distinguish between different signals carried by the same Smad given noisy chemical reactions. Since the number of TGF-family ligands are much larger than the number of receptor Smads, it is also not clear whether the cell can discriminate between signals carried by different Smads in the presence of significant cross-talk between them.

Other signaling pathways share a similar topology as the TGF – BMP – Smad pathway discussed here, such as the Jak-Stat pathway [6]. In fact they constitute what can be called the bow-tie network topology [7], wherein a large number of ligands activate a large number of genes through a smaller number of intermediary proteins. Thus cross-talk is hardwired into the structure of many mammalian signaling pathways.

In bacterial cells, a similar phosphorylation-based signal transduction motif is the two-component system. Here a cell surface receptor, usually a histidine kinase (HK), autophosphorylates when bound by a cognate ligand. The phosphate group is transferred to another protein called the response regulator (RR) which now becomes a transcription factor. One key difference between the bacterial and the mammalian systems is that the cell surface receptor in the latter is an enzyme for phosphorylation of the receptor Smad that carries the signal to the nucleus, and therefore one receptor molecule can phosphorylate many receptor Smads. In bacterial systems on the other hand, basically a single phosphate group is transferred, as in a relay race, from the cell surface receptor to the DNA. Bacterial systems also typically involve a smaller number of signaling proteins, i.e. their system size is smaller [8], [9].

Two component systems are found in nearly every bacteria and control myriad processes from nutrient sensing, chemotaxis, osmolarity control, quorum sensing and many others [10]–[12]. Most bacteria have many two component systems, and some are reported to have hundreds of them. Both the HK and the RR are paralogous gene families and they share significant amino acid and structural similarity within themselves [12]. It is possible therefore to imagine making use of cross-talk between pathways with similar structures to integrate signals into the final cellular decision. However despite a lot of research trying to look for examples of such cross-regulation, very few have been found [12]. The biochemical basis for cross-talk *in vivo* does exist with overexpression studies demonstrating that phosphotransfer between a HK and its noncognate RR is possible *in vivo*. However bacteria appear to use many methods to explicitly suppress cross-talk between two component systems. The known mechanisms of cross-talk suppression include: (i) bifunctional histidine kinases that act as a phosphatase for response regulators (ii) competition by the cognate RR that phosphotransfers with greater efficiency due to biochemical specificity and (iii) relatively low concentration of the HK to optimize the competition by the cognate RR [12].

There are also a few examples of situations where more than one HK signals through the same RR. For example, in the sporulation pathway of *B. subtilis*, four HK's can signal through a single response regulator, Spo0F [13]. Similarly, in the quorum sensing pathway of *V. harveyi*, three histidine kinases, LuxN, LuxQ and CqsS can phosphotransfer with the response regulator LuxU [14], [15]. These many-to-one branched pathways beg the question as to how bacteria can distinguish between signals originating from different HK's. The *V. harveyi* quorum sensing signal was studied in Ref. [15] which concluded that the bacteria could not distinguish between signals originating from the different HKs based on steady state values of a single phosphorylated RR alone. However the effects of cross-talk on the ability to distinguish between signals originating from different two-component systems have not yet been studied. This question gains significance given that bacteria appear to minimize cross-talk and do not appear to make use of it for cross-regulation [12].

To gain some insight into these issues, we turned to information theory. Information theory was developed in the late 1940s to ask abstract questions about general communication channels and has been used to gain insights about biological communication in the cell [16]–[18]. Information theory can be regarded as an application of probability theory to the problems of determining limits of information transmission in any communication channel, and it allows us to quantify the quantity of information that a network carries.

3.3 Methods

From the point of view of information theory, a signal transduction network that takes an extra-cellular signal and converts it into a concentration of a transcription factor is a noisy communication channel whose task is to convey information about the extra-cellular signals to the decoding and the decision making apparatus in the nucleus of the cell [17], [19]. If the distribution of the extra-cellular signal is given by a joint probability distribution function $p(X,Y)$, where X and Y are the levels of the extra-cellular signals, the total uncertainty of $p(X,Y)$ is measured by the Shannon entropy of their joint probability distribution function,

$$H(X,Y) = -\sum_{i,j} p(x_i,y_j) \log p(x_i,y_j).$$

Equation (1)

Here we follow the convention that the random variable is denoted by the capital letter, as in X, while the specific values it takes is the respective lower case letter, such as x_i. The information about the value of (X,Y) on the surface is encoded into the concentration of the output Z, which in our case is the concentration of an activated transcription factor. This is decoded by the genetic architecture and the appropriate response determined. We assume here that the cell has developed optimal decoding methods over millions of years of evolution and concentrate only on the information present in the output signal, Z. The information contained in Z about the value of (X,Y) can be thought of as the reduction in uncertainty about (X,Y) by knowledge of Z. This is measured by a quantity called the *mutual information* between (X,Y) and Z [20], denoted $I(X,Y;Z)$, which is given by,

$$I(X,Y;Z) = H(X,Y) + H(Z) - H(X,Y,Z).$$

Equation (2)

Now we can ask to what extent the cell can discriminate between the signals it receives from the two external ligands. Following Ref. [15], this is equivalent to asking how much the uncertainty in X is decreased by knowledge of Z, independent of the value of Y, and can be estimated by the mutual information between X and Z independently of Y, denoted $I(X;Z)$. A similar calculation can be performed for $I(Y;Z)$.

The mutual information is usually calculated using logarithms to base 2 and measured in bits. One bit corresponds to knowledge about the state of a 2-state system. The information content is therefore an absolute measure and can be given a physical meaning. In general, if the mutual information between X and Z is N bits, the cell should be able to distinguish up to 2^N distinct states of X from knowledge of Z, under the assumption of efficient decoding.

The physiological probability distribution function for the external input, the (X,Y) vector, is unknown. However since we are exploring the information processing capabilities of the networks in question, we can construct an arbitrary probability distribution function of the inputs. The simplest assumption is to start with a discrete distribution of (X,Y) with equal probabilities, i.e. a discrete uniform distribution over a two-dimensional range. In physiological conditions it is certainly likely that the cell needs to distinguish between coarsely positioned discrete values or ranges of the external ligands than very small differences (though the latter may be appropriate for some sensory cells), hence we chose a 26×26 grid of X and Y values spaced by 10 molecules from 0 to 250. The probability of seeing any of the combinations of (X,Y) is therefore,

$$p(x_i, y_j) = \frac{1}{676}.$$

Equation (3)

We keep this number fixed throughout this paper. This also sets the total uncertainty in the external distribution to be 9.4 bits. We use this number to calculate the efficiency of information transfer later in the paper. Note that this exercise is equivalent to performing an experiment where the cell is exposed to each of the 676 different combinations of the external ligands many times, and a histogram of responses constructed. Thus assuming a uniform distribution of the external ligands is the most appropriate assumption from the point of view of an *in vitro* experiment on the lines of Ref. [21].

In terms of probability distribution functions of the output and the input, the mutual information can be written as,

$$I(X,Y;Z)=\sum_k \sum_i \sum_j p(x_i,y_j,z_k)\log\left(\frac{p(x_i,y_j,z_k)}{p(z_k)p(x_i,y_j)}\right).$$

Equation (4)

$$I(X,Z)=\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,z_j)\log\left(\frac{p(x_i,z_j)}{p(z_j)p(x_i)}\right)$$

Equation (5)

where the joint probability distributions are defined in the usual way as,

$$p(x_i,y_j,z_k)=p(z_k|x_i,y_j)p(x_i,y_j)$$

Equation (6)

$$p(x_i,z_k)=\sum_j p(z_k|x_i,y_j)p(x_i,y_j)$$

Equation (7)

To estimate these probabilities, we perform stochastic simulations of the signal transduction network using the Gillespie algorithm. For each one of the possible 676 inputs we carry out 100 stochastic simulations of each network we consider using the Gillespie algorithm [22]. The Gillespie algorithm is an exact Monte Carlo simulation of the chemical Master equation that governs the stochastic evolution of the system. The models that we study are shown in Fig. 3.1 and are described as follows: (i) Fig. 3.1A shows the simplest model where two ligands operating through two surface receptors phosphorylate a single Smad. The output signal is the maximum accumulation of phosphorylated Smad. (ii) Fig. 3.1B shows the case where a protein called a Co-Smad binds to the activated Smad molecule. The signal at the nucleus then consists of a phosphorylated Smad and a heterodimer of a Smad with a Co-Smad, i.e. the output is bivariate. (iii) Fig. 3.1C shows the model with two Smads that are specific to the different receptors, and transduce the information to the nucleus. The output signal in this case are the maximum accumulations of the two phosphorylated Smads. (iv) Finally, Fig. 3.2 shows the network diagram of two bacterial two-component signaling systems. Here the receptor molecule, usually a histidine kinase, autophosphorylates on ligand binding, and the phosphate group is transferred to another protein called a response regulator. The output signal at the nucleus are the levels of the activated response regulator. The development of the Smad models is detailed in Appendix I: Table S3.1, Table S3.2 and Text S3.1. Development of the two-component model is detailed in Appendix I: Table S3.3 and Table S3.4.

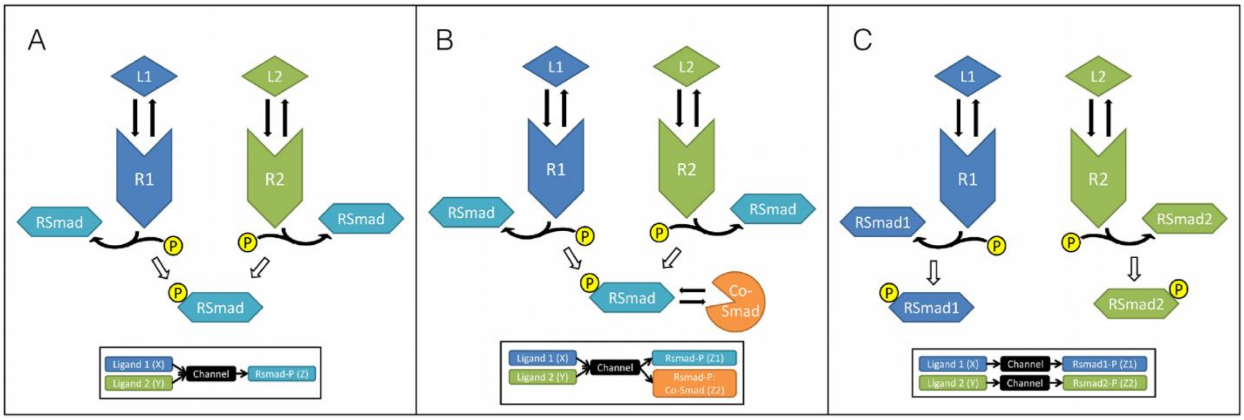


Figure 3.1. Smad signaling pathway network models.
(**Model A**) A single channel of one RSmad with a single output. (**Model B**) A single channel with two outputs, the phosphorylated RSmad and the RSmad:Co-Smad heterodimer. (**Model C**) The dual channel with two distinct RSmads and two outputs. The insets diagram the information transmission topology of signal (ligand), channel, and output (complexes). Note that this diagram represents a phosphorylation reaction by the receptors, not a phosphotransfer.

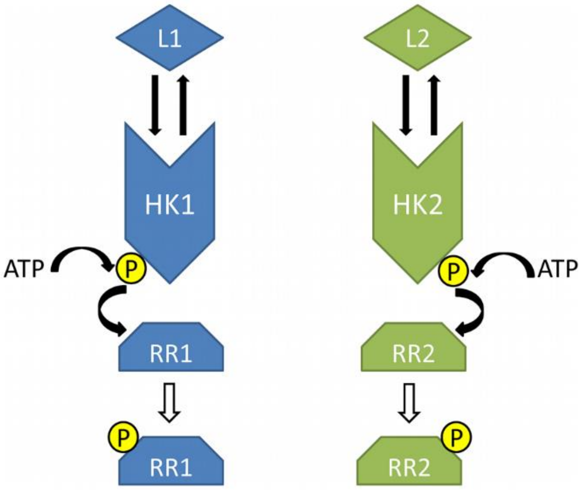


Figure 3.2. Bacterial two component system schematic model.
Note that unlike the mammalian system a single phosphate group is transferred from the cell surface histidine kinase receptor to the response regulator.

The parameters for the simulations are mainly taken from previously published work on Smad signaling and bacterial signaling and are listed and discussed in Appendix I: Table S3.2 and Table S3.4. The stochastic simulations allow us to construct a distribution of output concentrations by binning at specified times. Previous work on Smad signaling indicates that it is not the temporal pattern of Smad accumulation but the accumulation in the nucleus that is the relevant physiological concentration [23], [24]. For the simple Smad model of Fig. 3.1A we therefore choose the maximum accumulation of the activated proteins as the output variable, or the Z variable, and calculate the mean and the standard deviation of the maximum accumulation from the stochastic simulations. We then assume then that Z is normally distributed with the same mean and standard deviation. This is justified since the distribution of Z is the distribution of the mean of some other distribution, probably related to the extreme value distributions, and therefore by the Central Limit Theorem should be approximately normal.

The relevant distribution for each input combination is then binned to transform the normal distribution into a discrete distribution of Z-values. Since the information transfer naturally depends upon the bin size chosen to discretize Z, we decided to choose a bin size of 1. This is because due to discreteness of molecules this is the smallest relevant bin size. In effect we are assuming that the nucleus can make out differences of even one molecule of Z, which is undoubtedly an overestimate of cellular information processing quantities. Thus the values of the mutual information we calculate should be considered as upper bounds of the information transfer with uniformly distributed inputs. After the binning the conditional probability distribution of Z becomes:

$$P(z|x_i,y_i) = \int_{z-\Delta}^{z+\Delta} \frac{1}{\sqrt{2\pi\sigma_{x,y,z}^2}} \exp\left(-\frac{[z-\mu_{x,y,z}]^2}{2\sigma_{x,y,z}^2}\right) dz$$

Equation (8)

where $2\Delta = 1$ is the bin size. The values of the other probabilities required can be obtained from this equation by standard means using Eq. (7). These probabilities are then inserted into Eq. (4) and Eq. (5) to calculate the total and partial mutual information.

Note that the parameters are chosen so that the signal saturates above 250 molecules of each ligand, as shown in Appendix I: Fig. S3.1. Therefore the ligand concentration covers the dynamic range of the signaling network.

Incorporation of the Co-Smad as in Fig. 3.1B converts the output from a scalar into a vector, $\mathbf{Z} = (Z_1, Z_2)$, where Z_1 is the level of the activated Smad and Z_2 the level of the heterodimer. The probability distribution function of the output vector is therefore the joint probability distribution function of (Z_1, Z_2). In accordance with our previous assumption we assume that this is given by the appropriately binned bivariate normal distribution, i.e.

$$P(z_1, z_2 | x_i, x_j) = \int_{z_1 - \Delta}^{z_1 + \Delta} \int_{z_2 - \Delta}^{z_2 + \Delta} \frac{1}{\sqrt{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}} exp\left(- \frac{1}{2(1-\rho^2)} \left[\frac{(z_1 - \mu_{x,y,z_1})^2}{2\sigma_{x,y,z_1}^2} + \frac{(z_{x,y,z_2} - \mu_2)^2}{2\sigma_{x,y,z_2}^2} - \frac{2\rho(z_{x,y,z_1} - \mu_{x,y,z_1})(z_{x,y,z_2} - \mu_{x,y,z_2})}{\sigma_{x,y,z_1}\sigma_{x,y,z_2}} \right] \right)$$

Equation (9)

where $\mu_{x,y,z_1}, \sigma_{x,y,z_1}$ and $\mu_{x,y,z_2}, \sigma_{x,y,z_2}$ are the means and standard deviations of Z_1 and Z_2, ρ is the correlation coefficient and 2Δ is the bin size. Here we take the bin size to be 10 molecules to decrease the number of summations to be performed. As before the mean, standard deviation and correlation coefficients of the output is measured from the stochastic simulations. The mutual information between $\mathbf{Z}$ and the inputs (X, Y) jointly or singly is given by Eq. (4) and Eq. (5) as before with the joint probability distribution function of $\mathbf{Z}$ used in place of the univariate probability distribution function.

When we have two Smad proteins as in the model of Fig. 3.1C, we assume that the nucleus is only reading the levels of the phosphorylated RSmad and ignore dimerization, since our results show, as discussed later, that binding by the Co-Smad and dimerization or oligomerization are not likely to affect the information transfer. The input as before is the matrix of values (X,Y) and the output now is the level of two phosphorylated Smad proteins, (Z_1,Z_2). We again perform stochastic simulations to determine the mean and standard deviation and the correlation matrix of (Z_1,Z_2) and use those values and the bivariate normal distribution Eq. (9) above to calculate the probabilities of (Z_1,Z_2) lying in discrete bins. These probabilities are then used to calculate the mutual information between the input signal and the output signal.

Bacterial two-component systems were modeled following Ref. [8], and the parameter values were mostly taken from the same reference. The system is schematically shown in Fig. 3.2, the reactions and parameter values are detailed in Appendix I: Table S3.3 and Table S3.4 and the dynamic range of the signal at these parameter values is shown in Appendix I: Fig. S3.2. As before stochastic simulations of the reactions were carried out to determine the mean and the standard deviation of the signal, which is here taken to be the steady state value of the phosphorylated response regulator. The signal itself is assumed to be distributed according to a bivariate normal distribution as above (Eq. 9), and the information measures are calculated as before. All the calculations performed for the case when the output was bivariate used the same bin-size $2\Delta = 10$.

3.4 Results

3.4.1 Two ligands and a single Smad

We begin from the simplest possible model of Smad signaling shown in Fig. 3.1A. In this model, two extracellular ligands can bind to their cognate receptor heterodimer. The bound complex can then recruit and phosphorylate Smad proteins which then become transcription factors. We assume that each BMP ligand does not interact with the other receptor pair; in other words, there is perfect specificity at the receptor level, but both receptors signal through a single phosphorylated Smad protein. The detailed reactions and the parameter values chosen are shown in Appendix I: Table S3.1 and Table S3.2. We ignore the role of the Co-Smad and oligomerization of the Smads initially (see below).

The results of our calculations are shown in Fig. 3.3 and in Table 3.1. We find that this simple network, which we call Model A, is not efficient in information transfer from the external ligand concentration vector (X,Y) to the input Z. In fact as we show in Table 3.1, only about 3.6 bits of information about the vector (X,Y) are contained in Z, which corresponds to the ability to distinguish between only about 12 states of concentration values of the external ligand. This corresponds to about 40% information transfer efficiency about the external distribution of (X,Y).

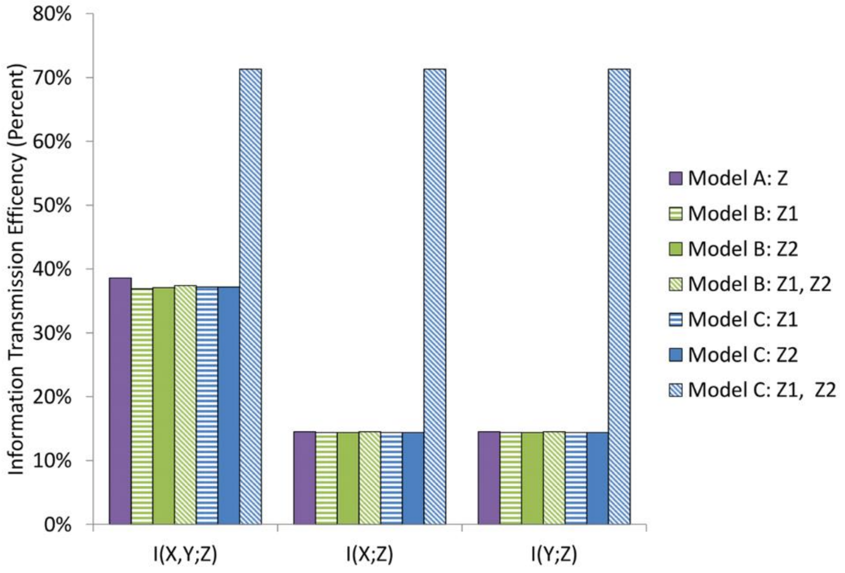
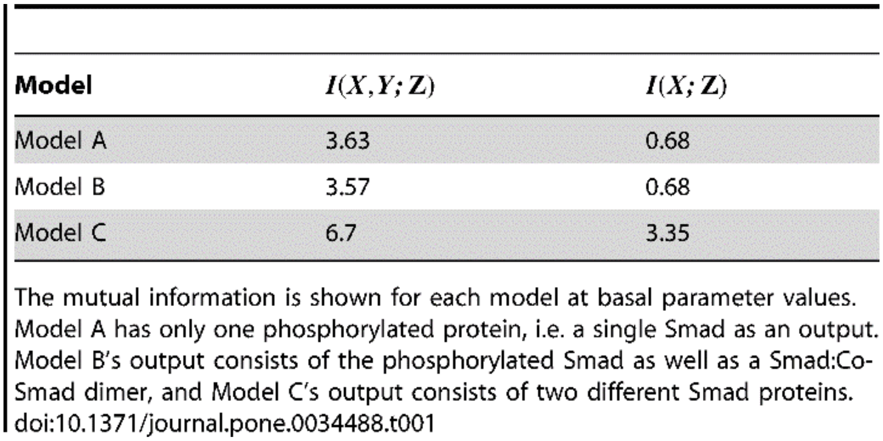


Figure 3.3. Summary of calculations of $I(X,Y;\mathbf{Z})$, $I(X;\mathbf{Z})$ and $I(Y;\mathbf{Z})$ for the three Smad pathway network topologies considered in this paper.

Information transfer efficiency as a percentage of the total uncertainly in the external distribution of ligands is shown in the y-axis. Model A is the model with a single Smad (Z), Model B is the model with a RSmad (Z_1) and a RSmad:Co-Smad heterodimer (Z_2) while Model C refers to the model with two RSmad proteins (Z_1 and Z_2). Information was calculated using either an individual output or both outputs (as is represented by Z1, Z2).

Table 3.1. Mutual Information Values.



Model	$I(X,Y;\mathbf{Z})$	$I(X;\mathbf{Z})$
Model A	3.63	0.68
Model B	3.57	0.68
Model C	6.7	3.35

The mutual information is shown for each model at basal parameter values. Model A has only one phosphorylated protein, i.e. a single Smad as an output. Model B's output consists of the phosphorylated Smad as well as a Smad:Co-Smad dimer, and Model C's output consists of two different Smad proteins. doi:10.1371/journal.pone.0034488.t001

However the ability of the cell to discriminate between X signals and Y signals is even poorer. At our basal parameter values we find that about 0.7 bits of information about X or Y alone is contained in the level of Z, which implies that the cell cannot even tell if X is high or low, since that requires one bit of information. This result is expected since both X and Y are activated in a completely symmetric way, so it is to be expected that the cell cannot distinguish between different levels of X when the effects of Y are potentially confounding. A possible way out for the cell to distinguish between ligands would be to increase the asymmetry in the kinetic responses elicited by the two ligands by, for example, making the phosphorylation rate of the RSmad by the receptor for X much higher than the other receptor. As shown in Fig. 3.4A we find that while this does lead to small increases in the information transmitted about X, it is at the cost of information about Y. Thus, maximizing information transfer about both stimuli is only possible when all rates are symmetric, i.e. at the cost of the ability to discriminate. This is the same as in two-component signal transduction systems in bacteria [15].

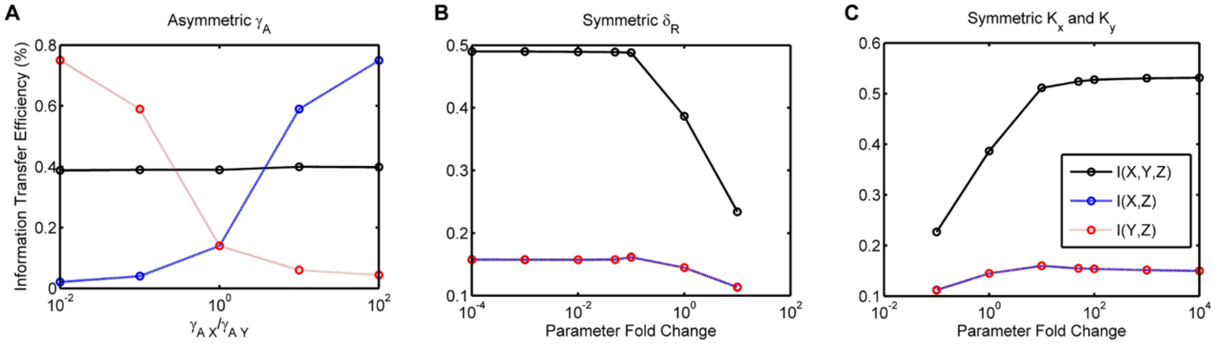


Figure 3.4. Parameter effects on information transmission for Model A.
(**A**) Information transfer efficiency (as a percentage of the total uncertainly in the external distribution of ligands) vs. the ratio of the rates of association of the RSmad to the X and Y receptors. (**B**) Information transfer efficiency vs. symmetrical increase/decrease of receptor degradation rate from the standard parameter rate. (**C**) Information transfer efficiency vs. symmetrical increase/decrease of ligand binding rate from the standard parameter rate.

Mutual information turned out to be rather insensitive to the parameter values that we choose for the simulation, as shown in Fig. 3.4B and C. Our parameter sensitivity analysis (Appendix I: Text S3.1, Fig. S3.3, Fig. S3.4, Table S3.5, Table S3.6) shows that most parameters had marginal effects on information transfer. The few parameters that could affect information transfer significantly are shown in Fig. 3.4. If the rate of receptor degradation is increased, it can decrease information transfer significantly, since the receptors degrade before a steady state binding equilibrium between the ligand and the receptors have been reached. However slowing the rate of degradation does not significantly increase information transfer, which plateaus at about 50% efficiency. Similarly, decreasing the equilibrium constant of binding between the ligand and the receptor leads to a decline in information transfer. However increasing the equilibrium constant beyond a point has no effect as information transfer again appears to plateau again at around 50%. Note that the rate of increase of information transfer is at best logarithmic.

What determines where the curve plateaus? The cell cannot really distinguish accurately between a signal from X and a signal from Y. The level of Z depends in fact on $(X+Y)$ since both ligands feed into the signaling machinery that determines the level of Z. Therefore the best that the cell, or any decoding algorithm can do is to distinguish between different levels of $(X+Y)$. The curve plateau is therefore related to the best possible discrimination between different levels of $(X+Y)$ that is possible at the parameter values of the simulation.

3.4.2 The Co-Smad does not increase information transfer

We then addressed the possible role of the Co-Smad in this network. In the biological network, the phosphorylated RSmad binds to a Smad protein called a Co-Smad and the heterodimer translocates to the nucleus and acts as a transcription factor. We wondered if the Co-Smad could help in translating small differences in the rate of phosphorylation of the RSmad by the two receptors into larger differences in the nucleus.

When we incorporate the Co-Smad (denoted as Model B), the output variable Z becomes a vector, $\mathbf{Z}=(Z_1,Z_2)$, where Z_1 is the level of phosphorylated Rsmad and Z_2 is the level of Rsmad:Co-Smad heterodimers. A diagram of this model is shown in Fig. 3.1B. Our calculations, summarized in Fig. 3.3 and Table 3.1, show that the coSmad heterodimer in fact does not contribute to the information transfer in the signaling network. This is not completely obvious since it could be imagined that at a given level of efficiency, adding Z_2 should increase total information transfer. As the data shows, efficiency at our basal parameter values is quite low, indicating that significant improvement is possible. However this cannot be achieved by adding a coSmad. The full details of this model can be found in Appendix I: Table S3.7.

Note that by the information processing inequality [20], information processing at an intermediate step in a Markov chain cannot increase the mutual information between the first step in the chain and the last. Therefore this inequality would predict that adding a Co-Smad should not be able to increase $I(X,Y;\mathbf{Z})$. However adding a Co-Smad cannot increase $I(X;\mathbf{Z})$ either since it acts symmetrically with respect to both channels since they transduce through a single Smad. Note that this implies that multimerization of the Co-Smad:RSmad complex cannot increase information transfer or signal discrimination either.

3.4.3 Multiple Smad pathways

We now ask what the effect would be if we had two Smad proteins instead of one. In other words, if each ligand had, along with a preferred receptor, a preferred Smad protein. A diagram of the model is shown in Fig. 3.1C, and the reactions are detailed in Appendix I: Table S3.1 and the parameter values in Appendix I: Table S3.2. We refer to this model as Model C. We assume as before that each ligand binds only to its cognate receptor. However now each receptor has a preferred Smad that it phosphorylates, which we assume is identical with the rate for the case of the single Smad. The catalytic rate by which each receptor phosphorylates its noncognate Smad protein can be varied. We call this rate the level of cross-talk between the two pathways.

When the level of cross-talk is zero, each ligand has its own dedicated Smad protein. Therefore, as expected, the total information transferred approximately doubles, at base parameter values, to about 6.7 bits (see Fig. 3.3 and Table 3.1). That is equivalent to the capacity to distinguish between about 104 states of the input signal (X,Y), which is quite a large number of states. The absolute efficiency of information transfer at basal parameter values has now reached a respectable value, and is about 71%. These results indicate that it is quite possible for signaling transduction networks to respond to relatively small changes in the levels of external ligands, and distinguish between many different states of these ligands merely by increasing the number of output proteins.

The ability to discriminate is as before measured by the mutual information between the output (Z_1,Z_2) and the input signal X (or Y) by summing up over all Y (or X). As shown in Fig. 3.3, we find a significant increase in the ability to discriminate with the mutual information $I(X;\mathbf{Z})=3.35$ bits at basal parameter values, which corresponds to about 10 states of the ligand concentration X. By using two Smads the cell has also restored the symmetry between the cell's ability to distinguish different levels of X and different levels of (X,Y) since the latter is approximately corresponds to 10 values of X and 10 values of Y, i.e. a total of 10^2 levels of (X,Y).

3.4.4 Cross-talk between Smad pathways

The above results are based on our calculations when the level of cross-talk is zero, i.e. each receptor talks with only its own cognate Smad. However most biological signaling pathways with multiple proteins usually have some cross-talk between proteins. Cross-talk between different proteins can be expected to decrease the efficiency of the information transmitted. To test what happens when the level of cross-talk increases, we then let each receptor phosphorylate the noncognate Smad at a fraction of the rate at which it phosphorylates its cognate Smad. This is implemented by changing the on-rate of binding of the non-cognate Smad with its non-cognate receptor from zero to some positive value, while the catalytic phosphorylation rate remains the same for all the cognate and non-cognate pairs. The ratio between the binding on-rate of the non-cognate pair with that of the cognate pair is thus a measure of the level of cross-talk, which can be varied both symmetrically, i.e. each receptor has the same amount of cross-talk as the other, or asymmetrically.

We find surprisingly that a significant level of cross-talk is tolerated before the information transmission efficiency decreases. As we show in Fig. 3.5, even when the effective phosphorylation of each receptor with the non-cognate Smad is 70% what it is with its cognate Smad, the total mutual information $I(X,Y;\mathbf{Z})$ as well as the partial mutual information $I(Y;\mathbf{Z})$ only marginally decreases compared to the case with no cross-talk. A significant decrease in the capacity of the channel requires that the cross-talk is greater than 80%. When the cross-talk is 100%, then as expected, both the Smad proteins are effectively the same, and the cell cannot do better than with a single channel. We find in fact that for total mutual information, it does a little worse, probably due to interference between the two pathways.

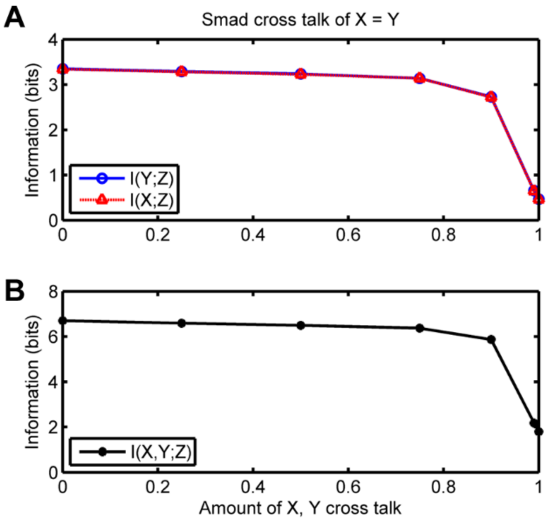


Figure 3.5. Information in bits vs fraction of cross-talk for the Smad Model C, with equal cross talk.
Note that cross-talk is defined as the ratio of the on-rate of a Smad protein for the non-cognate receptor to the on-rate for its cognate receptor. When cross-talk is zero, only the cognate receptor can phosphorylate the Smad; when cross-talk is one, both receptors are equally efficient in phosphorylation of that Smad. In this plot the cross-talk between the receptor for X and output Z_2 is the same as that between the receptor for Y with output. (**A**) Partial mutual information. (**B**) Total mutual information.

In Fig. 3.6 we show what happens when the cross-talk between one receptor-non-cognate Smad pair is kept fixed at either zero or one while the other varies. Here we see that if one receptor does not talk at all to its non-cognate Smad, it does not matter even if the cross-talk of the other receptor for the non-cognate Smad is 1; the mutual information is completely unaffected. If on the other hand the cross-talk between one receptor-non-cognate Smad pair is kept at 1, increasing the cross-talk of the other pair up from zero begins to adversely affect the information content of the channel only when the cross-talk crosses about 70%. Therefore information content suffers only when the cross-talk efficiencies are symmetrically high.

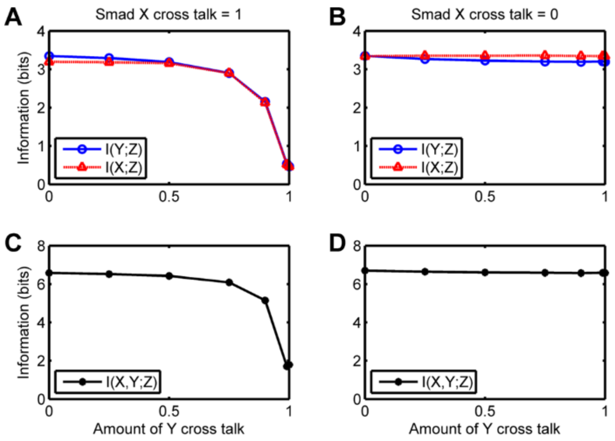


Figure 3.6. Mutual information as a function of cross-talk for Smad Model C when the cross-talk is varied asymmetrically.
(**A** and **C**) Cross talk for the receptor for X with its non-cognate Smad held at 1. (**B** and **D**) Cross talk for the receptor for X with its non-cognate Smad held at 0. (**A** and **B**) Partial mutual information. (**C** and **D**) Total mutual information.

This scenario has some interesting implications for protein evolution and information transfer. Due to this relation between cross-talk and $I(X,Y;Z)$ for both the symmetrical and the asymmetrical cases discussed above, there does not appear to be a strong tendency for minimization of cross-talk on the basis of information transfer alone. However if signaling is relatively robust against high levels of cross-talk, it is robust against having overlapping or partially redundant pathways. Redundancy has many protective advantages in biology, and in mammalian signaling for example, partially redundant pathways can compensate for defects in other pathways [25]. It also becomes possible to imagine the development of new functionalities from small mutations in signaling proteins as well as the development of cross-regulation wherein cross-talk is exploited to integrate signals coming from many external stimuli.

3.4.5 Cross-talk in bacterial two component systems

We now turn to bacterial two component systems. The basic structure of a bacterial two component system is shown in Fig. 3.2. This consists of a cell surface receptor, usually a histidine kinase (HK) that can autophosphorylate when bound with its cognate ligand. The phosphate group can then be transferred to another protein molecule, generically called a response regulator (RR). The phosphorylated RR then turns on specific genes in the bacterial DNA [10], [11].

The main differences between the mammalian system and the bacterial system are the system size and the difference in the method of enzymatic activity i.e. the receptor molecules in bacteria autophosphorylate followed by a phosphotransfer to the response regulator. Mammalian cells on the other hand have receptor molecules that phosphorylate the cognate signaling protein, transferring a phosphate group usually present in excess in solution to the protein in question. It turns out that these differences do not necessarily lead to a change in total information transfer in the absence of cross-talk. Our calculations based on parameter values from [8] and ligand concentrations that almost cover the dynamic range of the system response show that two separate response regulators can transduce, in the absence of cross-talk, about 6.9 bits of information when taken together, which is about the same as the Smad system. The bacterial system size with these parameter values is about an order of magnitude smaller than the Smad system size as shown in Appendix I: Fig. S3.2. This is consistent with measured protein concentrations in many two component systems [8], [9].

However when cross-talk is added to the system it shows a very different behavior. The mutual information $I(X,Y;\mathbf{Z})$, between the external ligands (X,Y) and the level of phosphorylated response regulators $\mathbf{Z}=(Z_1,Z_2)$ begins declining monotonically as cross-talk between the two HK's is symmetrically increased from zero as shown in Fig. 3.7. The mutual information between one external ligand and the output $\mathbf{Z}$ also declines in a similar manner. This is in sharp contrast with the behavior of the mammalian system as discussed above. The bacterial cell is more robust to cross-talk when it is only one-sided, i.e. only one HK can phosphotransfer to both RRs. In this case, as we see in Fig. 3.8A the decline in $I(X,Y;\mathbf{Z})$ and the decline in $I(X;\mathbf{Z})$ is much slower. However if one HK is already promiscuous then increasing the cross-talk of the other leads to an even sharper decline in both total mutual information as well as partial mutual information (Fig. 3.8B).

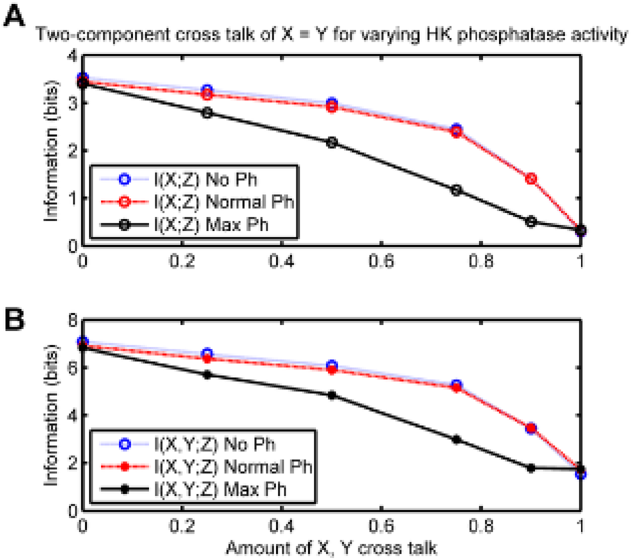


Figure 3.7. Information in bits vs. cross-talk for the two-component model, with equal cross talk and various HK phosphatase activities.
(**A**) Partial mutual information and (**B**) total mutual information. The red line shows the default phosphatase activity where strength of phosphatase activity toward the non-cognate RR varies with level of cross talk. The blue line shows a system where the HK has no phosphatase activity. The black line shows where the strength of phosphatase activity toward the non-cognate RR is maximum regardless of the level of cross-talk.

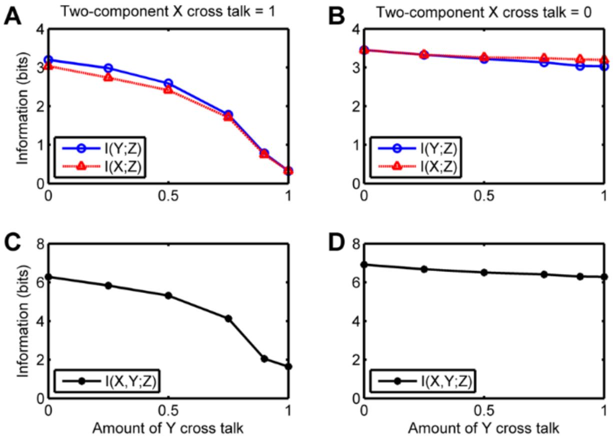


Figure 3.8. Information in bits vs. cross-talk for the two-component model, with asymmetric cross talk.
(**A** and **C**) Cross talk for the receptor for X with its non-cognate Smad held at 1. (**B** and **D**) Cross talk for the receptor for X with its non-cognate Smad held at 0. (**A** and **B**) Partial mutual information. (**C** and **D**) Total mutual information.

Modeling studies have argued that phosphatase activity of a HK with respect to its RR can buffer the system against cross-talk by dephosphorylation of weak signals from a non-cognate HK. However this method is probably unlikely to be very efficient when both external ligands are present and therefore both HKs are being activated. We tested this by simulating the system when the phosphatase activity of the HK was kept at a maximum regardless of cross-talk (black line), when the phosphatase activity for the non-cognate RR varies proportionately with level of cross-talk (red line), and when the HKs have no phosphatase activity (blue line) Fig. 3.7. As shown in Fig. 3.7 the phosphatase activity of the HK has no impact on cross-talk when both external ligands are present and the mutual information measures $I(X,Y;\mathbf{Z})$ and $I(X;\mathbf{Z})$ decrease monotonically. In the case of maximum phosphatase activity a sharper decline is seen, which may be a consequence of suppression of the signal to the cognate RR due to the high phosphatase activity.

In order to understand whether the degradation of information content was due to the difference in system size or due to the kinetic differences between the two pathways, we took the two-component model and changed parameters (Appendix I: Table S3.8 and Fig. S3.9) until we obtained a dynamic range that was approximately equivalent to the Smad model. Similarly, we took the Smad signaling model and changed parameters to obtain a model that yielded protein numbers that were of the same order as that of the two-component model (Appendix I: Fig. S3.10). Results of the two large-protein-number models are are shown in Fig. 3.9A, and they indicate that in fact at high protein numbers the two modes of signal transduction are identical. The results from the two small-protein-number models, shown in Fig. 3.9B, suggest that at small protein numbers there is still some difference between the two cases, that could be due to the small remaining difference in protein numbers, the higher level of noise of the two-component circuit, or the mode of receptor action.

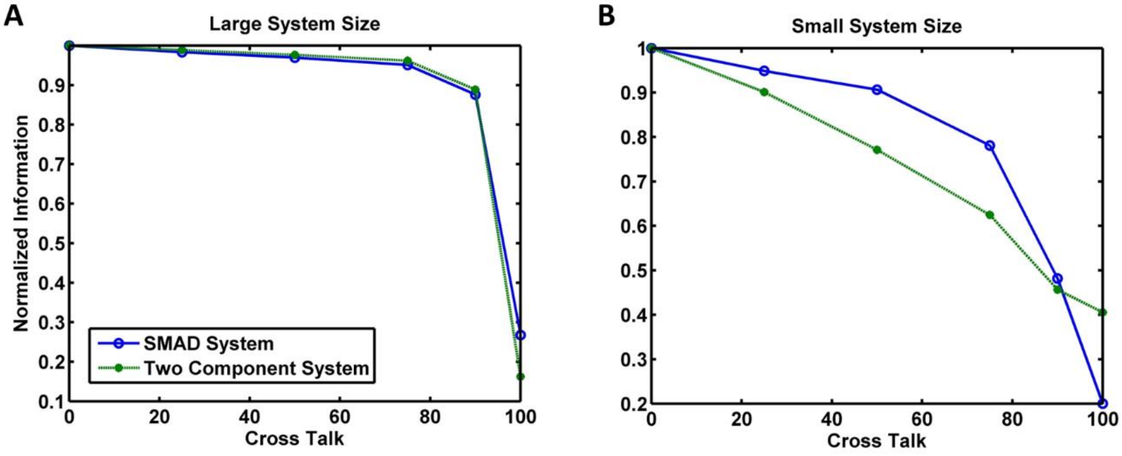


Figure 3.9. Information in bits vs. symmetrically varying cross-talk, a comparison of the effects of system size and the mode of phosphotransfer.
(A) The Smad system and the two-component system for large system sizes (B) the Smad systems and the two-component system for small system sizes.

Why do the smaller system sizes that we simulate in this work show a greater degradation of information with increasing cross-talk? Smaller protein numbers are associated with larger relative fluctuations due to the intrinsic stochasticity of signaling networks. Symmetric crosstalk not only increases the total noise in the system, it also leads to decrease in the absolute number of useful molecules for each signaling channel, thereby further decreasing the signal to noise ratio of each channel. This could very well be the reason why we see increasing sensitivity to cross-talk with a smaller system size.

The monotonic decline in information transfer with increasing cross-talk seen in our calculations suggest that small signaling systems, such as those characteristic of some bacterial two-component networks, cannot function efficiently in the presence of cross-talk without increasing the number of signaling proteins by an order of magnitude or so. Energetically it is cheaper to use two independent signaling pathways for transducing information, as they can transfer as much information with a smaller number of proteins. This suggests that evolution should have led to two component systems evolving to be relatively insulated from each other, as cross-talk would lead to a decline in fitness. This could be one reason why we do not find much cross-regulation between different two-component pathways. We can also predict, based on these arguments, that if cross-talk is introduced between two two-component pathways (by say a lateral gene transfer event), we should initially see a decline in fitness, and evolution should eventually drive the system to eliminate cross-talk between these pathways altogether.

3.5 Discussion

We have used information theoretic methods to study the transmission of information in simple signaling networks based on the Smad signaling pathway of the TGF-β proteins in mammalian cells and two-component systems in bacteria. It is often assumed that what is important in gene circuits or the cell in general is bistability, i.e. having two states – ‘on’ and ‘off’. However in principle the information transmitted by a simple signaling pathway like the

Smad signaling pathway can allow the cell to perform much more sophisticated information processing than simple binary decisions [26].

It is not clear whether signal transduction networks in cells actually transduce more than one bit of information. Recent experimental studies on Tumor Necrosis Factor Alpha signaling have claimed that only one bit of information is carried in several important signaling networks [21]. However in principle many signal transduction networks appear to have the ability to distinguish more than binary levels of extra-cellular signals. Some sensory cells like neurons and hair cells in the ears have extremely accurate sensing capabilities, that have been optimized over millions of years of evolution. It has therefore also been argued that evolution should have optimized the information transmission capabilities of signal transduction networks [17]. In this paper however we do not use the optimality assumption but rather ask whether the information transmitted to the nucleus could potentially allow the cell to reconstruct the distribution of the external signals that led to the signal. In particular we ask whether these signaling pathways have the capacity to allow the cell to distinguish between signals received by two external ligands.

Our results are based on calculations of two measures, the total mutual information $I(X,Y;\mathbf{Z})$ and the partial mutual information $I(X;\mathbf{Z})$. The total mutual information $I(X,Y;\mathbf{Z})$ tells us the maximum number of states of the external ligand concentration (X,Y) that the cell could in principle distinguish, assuming efficient decoding mechanisms exist. Similarly, $I(X;\mathbf{Z})$ tells us the number of states of the ligand concentration X that the cell could distinguish from knowledge of $\mathbf{Z}$ alone. Both of these measures depend upon parameter values and concentrations, as well as upon the topology of the network. In this study we assume reasonable parameters and calculate the information transmission measures at these parameter values. We then vary all parameters by large amounts to see whether the qualitative results are sensitive to the choice of parameter values. We find that our qualitative results are very robust against wide variations in most parameter values. Parameters adjusted are shown in Appendix I: Table S3.5, Table S3.6 and Table S3.7 and the results are shown in Fig. S3.5, Fig. S3.6, Fig. S3.7 and Fig. S3.8.

We find, in agreement with previous results [15] that the cell cannot distinguish between different levels of the external ligands $\boldsymbol{X}$ or $\boldsymbol{Y}$ based on the level of phosphorylated Smad protein if the receptors for the two external ligands are symmetric in terms of their effective rates of phosphorylation of the Smads. While some specificity can be introduced by making the receptors asymmetric, this is at the cost of one of the two external signals. The ability to discriminate is not helped by addition of a Co-Smad to the system.

However we find that addition of another output protein, i.e. another R-Smad, increases both the total information carried as well as dramatically increases the cell's ability to distinguish between different levels of the external ligands. While in the case of a single Smad, the cell could not distinguish even between high and low levels of a single external ligand, with two Smads the cell can, in principle, distinguish between 10 different levels. The multiplicity of signaling proteins that carry information to the nucleus in pathways like the Smad signaling pathway are probably a direct consequence of this dramatic increase in information transmission.

It should be expected that as the cross-talk between the receptors of the two output proteins, (Z_1, Z_2) increases, it leads to a decrease in the ability of the cell to discriminate and in the total information carried by the channel. When cross-talk is 100% in both directions, effectively both Z_1 and Z_2 are indistinguishable from each other and we find, as expected, that no additional information is carried by the communication channel compared to a single Z protein. However as the level of cross-talk is decreased below 100%, we find a relatively steep increase in both measures that almost reach a plateau by the time the cross-talk drops to below 70%. In other words we find that contrary to intuition, a high level of cross-talk is not very deleterious to information transmission by the Smad pathway, or other similar pathways, in mammalian cells.

This result has potentially significant implications. Consider the situation where a single signaling pathway is altered by a heterozygous mutation in one allele of the gene corresponding to a Smad-like protein. If the heterozygous mutation is in an important residue and it leads to one of these proteins becoming preferred for a previously existing function, or acquiring a new function, it would result in a significant increase in information transfer, possibly conferring an evolutionary advantage that could lead to the mutation being fixed in the population. For example sequence analysis shows that the human Smad proteins cluster into two groups, one associated with BMP signaling and the other associated with TGF-β signaling [27] and both clusters share significant sequence similarity. It is possible therefore that each cluster arose by mutations in a single protein that was beneficial because of the resulting increase in information transfer despite the high level of cross-talk. Similarly BMP2 and BMP4 share 92% sequence similarity but play some non-redundant roles in cellular signaling. It is possible therefore that BMP 2 and 4 could have originated by mutations in a single BMP protein that created ‘new’ extra-cellular ligands with different receptor specificity and <100% cross-talk with each other, leading to a significant increase in information transmission. Increases in information transmission due to such mutations could be one of the important sources of positive selection of mutations in signal transduction.

Another very common scenario is duplicate genes that are ubiquitous in human and other genomes. About 15% of the human genome consists of duplicate genes, many of which have diverged in function [28]. The creation of a duplicate gene would pave the way for gradual divergence of each gene [25]. The acquisition of new functions again would be crucially helped by the fact that the cell can deconstruct signals coming from each protein despite cross-talk. It is possible that signaling pathways depending upon closely related sets of genes diverged from each other due to such processes. As long as the cross-talk between these pathways is not close to one, there are no deleterious effects on the original pathway. Furthermore, the existence of overlapping pathways does provide protection due to development of redundancy in the cell, and leads to the possibility of cross-regulation, i.e. integration of multiple signals into the same decision process [25]. These results are not exclusive to the Smad pathway as there are a number of mammalian pathways with similar topology, such as the Jak-Stat pathway [6], where robustness against cross-talk when surface receptors are efficient kinases for transcription factor molecules may have played a role in the development of complexity in signaling networks.

The dominant cause of the relative insensitivity of the system towards increasing cross-talk appears to be the system size. In smaller systems such as bacterial two-component systems, we see an almost monotonic decline of total mutual information and partial mutual information when cross-talk exists between two HKs for their non-cognate RRs. The sensitivity to cross-talk in smaller two-component systems may be one reason why bacteria, who can have hundreds of such systems, expend considerable effort to avoid cross-talk and keep them insulated from each other. It is interesting to note that many researchers had assumed that two-component signaling should naturally allow for cross-regulation between different pathways; however despite significant efforts, few examples of cross-regulation have been found [12]. Cross-regulation is not possible between systems where interference between two pathways leads to attenuation of information transfer. Our calculations would therefore predict that if cross-talk were introduced in a bacterium due to either lateral gene transfer or artificially, evolution would again tend to minimize the cross-talk between these two systems in order to overcome the fitness loss due to aberrant information transfer. Of course some bacterial signaling systems also involve thousands of proteins and are therefore large in the sense implied in this paper. Our analysis would predict that these larger systems are more likely to be insensitive to cross-talk, or to exploit it, compared to the smaller two-component systems.

We have not studied the effect that different input distributions may have on cross-talk between related signaling pathways, though we believe that they are unlikely to change our qualitative results. This is in accordance with Mehta et. al. [15] who found that different distributions of the ligand did not affect their results for a single transcription factor. It is possible that the efficiency of information transfer increases when the distribution of the extra-cellular ligand is different from the uniform distribution. The uniform distribution also has the maximum amount of uncertainty. Our results however easily translate into an experiment where a cell is exposed to different concentrations of two extracellular ligands repeatedly and the levels of the activated transcription factor measured. The histogram of these levels for each input combination is precisely the conditional probability distribution, $p(z|x_i,y_j)$.

The dependence of our results on system size may be because smaller systems have a higher noise to signal ratio due to the intrinsic stochasticity of chemical reactions. We are currently studying this relationship with the aim of uncovering a more precise quantitative relation between system size and the effect of cross-talk. Further work is also needed to understand how gene transcription networks can interpret signals coming from systems with an innately high level of cross-talk. Moreover, in future work we also need to understand what happens when there is cross-talk between more than two pathways at the same time. This is particularly relevant for TGF-β signaling and BMP signaling, since both of them have at least three Smad homolog's that are involved in information transfer from the receptor to the nucleus. Finally, our analysis also leads to the design of experiments to be performed that can confirm or falsify our predictions and uncover how cells make sense of the world in the presence of cross-talk.

REFERENCES

[1] Shi Y, Massagu J (2003) Mechanisms of tgf-beta signaling from cell membrane to the nucleus. Cell 113: 685–700.

[2] Marom B, Heining E, Knaus P, Henis YI (2011) Formation of stable homomeric and transient heteromeric bmp receptor complexes regulates smad signaling. J Biol Chem.

[3] Cho TJ, Gerstenfeld LC, Einhorn TA (2002) Differential temporal expression of members of the transforming growth factor beta superfamily during murine fracture healing. J Bone Miner Res 17: 513–520.

[4] Bais M, McLean J, Sebastiani P, Young M, Wigner N, et al. (2009) Transcriptional analysis of fracture healing and the induction of embryonic stem cell-related genes. PLoS One 4: e5393.

[5] Lavery K, Swain P, Falb D, Alaoui-Ismaili MH (2008) Bmp-2/4 and bmp-6/7 differentially utilize cell surface receptors to induce osteoblastic differentiation of human bone marrow-derived mesenchymal stem cells. J Biol Chem 283: 20948–20958.

[6] Kisseleva T, Bhattacharya S, Braunstein J, Schindler CW (2002) Signaling through the jak/stat pathway, recent advances and future challenges. Gene 285: 1–24.

[7] Ma'ayan A (2009) Insights into the organization of biochemical regulatory networks using graph theory analyses. J Biol Chem 284: 5451–5455.

[8] Igoshin OA, Alves R, Savageau MA (2008) Hysteretic and graded responses in bacterial two-component signal transduction. Mol Microbiol 68: 1196–1215.

[9] Batchelor E, Goulian M (2003) Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. Proc Natl Acad Sci U S A 100: 691–696.

[10] Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. Annu Rev Biochem 69: 183–215.

[11] Robinson VL, Buckler DR, Stock AM (2000) A tale of two components: a novel kinase and a regulatory switch. Nat Struct Biol 7: 626–633.

[12] Laub MT, Goulian M (2007) Specificity in two-component signal transduction pathways. Annu Rev Genet 41: 121–145.

[13] Jiang M, Shao W, Perego M, Hoch JA (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in bacillus subtilis. Mol Microbiol 38: 535–542.

[14] Henke JM, Bassler BL (2004) Three parallel quorum-sensing systems regulate gene expression in vibrio harveyi. J Bacteriol 186: 6902–6914.

[15] Mehta P, Goyal S, Long T, Bassler BL, Wingreen NS (2009) Information processing and signal integration in bacterial quorum sensing. Mol Syst Biol 5: 325.

[16] Bialek W, Setayeshgar S (2005) Physical limits to biochemical signaling. Proc Natl Acad Sci U S A 102: 10040–10045.

[17] Tkacik G, Callan CG, Bialek W (2008) Information flow and optimization in transcriptional regulation. Proc Natl Acad Sci USA 105: 12265–70.

[18] Lestas I, Vinnicombe G, Paulsson J (2010) Fundamental limits on the suppression of molecular fluctuations. Nature 467: 174–178.

[19] Tkacik G, Callan CG, Bialek W (2008) Information capacity of genetic regulatory elements. Phys Rev E Stat Nonlin Soft Matter Phys 78: 011910.

[20] Cover TM, Thomas JA (1991) Elements of Information Theory. Wiley Interscience.

[21] Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A (2011) Information transduction capacity of noisy biochemical signaling networks. Science 334: 354–358.

[22] Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry 81: 2340–2361.

[23] Hill CS (2009) Nucleocytoplasmic shuttling of smad proteins. Cell Res 19: 36–46.

[24] Clarke DC, Betterton MD, Liu X (2006) Systems theory of smad signalling. Syst Biol (Stevenage) 153: 412–24.

[25] Bomblies K (2010) Evolution: redundancy as an opportunity for innovation. Current Biology 20: R320–2.

[26] Ziv E, Nemenman I, Wiggins CH (2007) Optimal signal processing in small stochastic biochemical networks. PLoS One 2: e1077.

[27] Newfeld SJ, Wisotzkey RG (2006) Molecular evolution of smad proteins. Heldin P C & ten Dijke, editor, Smad Signal Transduction, Springer. pp. 15–35.

[28] Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3: 827–837.

CHAPTER 4: LOADS BIAS GENETIC AND SIGNALING SWITCHES IN SYNTHETIC AND NATURAL SYSTEMS [2]

4.1 Summary

Biological protein interactions networks such as signal transduction or gene transcription networks are often treated as modular, allowing motifs to be analyzed in isolation from the rest of the network. Modularity is also a key assumption in synthetic biology, where it is similarly expected that when network motifs are combined together, they do not lose their essential characteristics. However, the interactions that a network module has with downstream elements change the dynamical equations describing the upstream module and thus may change the dynamic and static properties of the upstream circuit even without explicit feedback. In this work we analyze the behavior of a ubiquitous motif in gene transcription and signal transduction circuits: the switch. We show that adding an additional downstream component to the simple genetic toggle switch changes its dynamical properties by changing the underlying potential energy landscape, and skewing it in favor of the unloaded side, and in some situations adding loads to the genetic switch can also abrogate bistable behavior. We find that an additional positive feedback motif found in naturally occurring toggle switches could tune the potential energy landscape in a desirable manner. We also analyze autocatalytic signal transduction

[2] The work presented in this chapter and in Appendix II has been published in PLoS Computational Biology and is reproduced here under the Creative Commons License. Lyons SM, Xu W, Medford J, Prasad A (2014) Loads Bias Genetic and Signaling Switches in Synthetic and Natural Systems. PLoS Comput Biol 10(3): e1003533. I am first author on the publication. The work is being presented here in its entirety to maintain the intellectual coherence of the project.

switches and show that a ubiquitous positive feedback switch can lose its switch-like properties when connected to a downstream load. Our analysis underscores the necessity of incorporating the effects of downstream components when understanding the physics of biochemical network motifs, and raises the question as to how these effects are managed in real biological systems. This analysis is particularly important when scaling synthetic networks to more complex organisms.

4.2 Introduction

A longstanding question about signal transduction and gene transcription networks is how modular are they. Here modularity means relative insulation of small subgraphs or motifs of the main network from each other [1]. This question is especially relevant for synthetic biology that aims to build artificial circuits from the bottom up [2]. It is also relevant for molecular biologists that aim to arrive at a quantitative understanding of a cellular decision, by, for example, isolating a crucial network module [3].

For synthetic biologists the challenge is now to move from simple network motifs such as pulse generators [4], genetic switches [5]–[8], logic gates [9], [10], and oscillators [11]–[13] to more complicated networks combining multiple motifs and networks in more complex organisms. Novel applications currently being explored include plant biosensors [14], hazardous waste remediation [15], clean fuel technology [16], and numerous medical applications [17]–[20]. Synthetic biologists hope to utilize biological modules in a manner similar to electrical circuit board components – plugging them together to attain a specific, and novel, function [21]. At the core of the concept of either breaking down complex biological systems into small modules, or even building complex systems from modules, is the belief that these modules will behave predictably in isolation and in connection. Recent theoretical and experimental work however [22]–[25] suggests that the functioning of modules may not be independent of the downstream components that they are connected to. Adding an additional binding reaction to the output of a gene regulatory network (or loading the network) may decrease system bandwidth [24] and substrate sequestration in covalent modification cycles may result in signaling delay [26]. *In vitro* studies find that there is significant load-induced modulation of the upstream module in an enzymatic signal transduction cascades [24]. Theoretical analysis has also shown that a load can change the fundamental properties of an oscillating circuit [27]. Thus understanding the effects of adding a load to the output of these technologically important network modules is required for a thorough understanding of the challenges of scaling up synthetic networks to higher levels of complexity.

Loads could also have noteworthy unrecognized effects in natural systems. In fact all natural systems have loads in some ways or the other. Motifs in signal transduction networks are connected directly to a transcriptional response, or to downstream proteins that may function as transcription factors or go on to activate transcription factors. Motifs in gene transcription networks have transcriptional outputs with protein domains that bind nonspecifically and specifically to binding sites on the DNA, apart from interacting with other transcription factors.

Circuits that function as switches play an important role in all biological signaling and gene transcription networks because they encode decisions. This change of state can be brought about by an external signal, or an internal accumulation of a protein, which can drive the system to a different steady state. Examples are the regulatory circuits for the cell cycle in yeast [28], mitogen-activated protein kinase cascades in animal cells [29]–[31], and the lysis-lysogeny switch in the λ phage [32]. Since many small circuits can show this kind of behavior, switches are among the earliest and most well studied of protein interaction circuits [33]. The genetic toggle switch, which was one of the first two synthetic circuits constructed, is a well-known synthetic example [5]. Given the ubiquity and importance of switch-like motifs, it is important to understand how their function could be affected by binding downstream partners.

These reasons prompted our theoretical study of the behavior of a simple genetic toggle switch [5], a toggle switch with positive feedback as well as a common positive-feedback based switch involving Ras activation in lymphocytes [29], [30] under a load on either one or both of its outputs. These circuits are shown in Fig. 4.1 and described below. The simple toggle switch is a widely studied and emulated synthetic network motif based on the mutual repression of two repressor proteins. However, naturally occurring toggle switches are often found connected to an additional positive autoregulatory component. For example in the competence system in B. subtilis, ComK represses the production of Rok and Rok represses the production of ComK; however ComK also has a strong positive feedback upon its own production [34]. Another example is found in the apoptosis network of many multicellular organisms, including mammals. Within the pathway controlling intrinsic apoptosis is a set of genes with double-negative repression, Casp3 and XIAP, again accompanied by positive autoregulation of Casp3 [35].

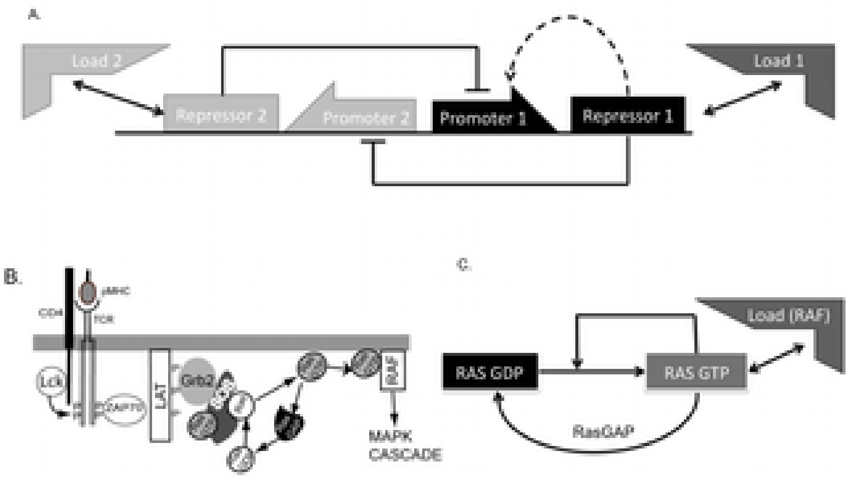


Figure 4.1. Schematic diagram of the circuits studied in this paper.
(A). The basic toggle switch is the network shown without the dotted line. Repressor 1 represses the production of Repressor 2 and vice versa. The dotted line denotes a positive feedback motif found in some natural circuits. (B). A cartoon of part of the MAPK activation pathway in T lymphocytes, adapted from [29], showing the role of Ras activation. Signals from peptide-MHC complexes are received at the TCR and lead to phosphorylation of the cytoplasmic chains of the TCR by the Src kinase, Lck. This recruits the kinase ZAP70 which trans-autoactivates and phosphorylates a scaffold called LAT, which recruits Grb2 and SOS to the plasma membrane. SOS activates Ras as as shown. (C) A simplified model of the Ras switch. RasGDP transforms into RasGTP via the enzyme SOS. However the catalytic rate of SOS increases when bound to RasGTP. This sets up an autocatalytic positive feedback. RasGTP is deactivated by enzymes called RasGAP's (among others).

The Ras protein is a G-protein found on mammalian cellular membranes that is important in many cellular processes and is an upstream activator of the MAPK pathway. Ras goes from a GDP-bound inactive form to a GTP-bound active form, often in a digital manner [30], and previous studies in lymphocytes have shown that RasGDP is activated to RasGTP via a bistable switch that arises from a positive feedback loop on its own activation via SOS (Son of Sevenless) [30]. However the Ras switch very naturally has an associated load, since to transduce the cellular signals down along the MAPK/ERK pathway, RasGTP naturally binds to Raf kinase. Thus the Ras switch system contains all the elements we need to study the effects of adding loads to a bistable switch which is based on a positive feedback loop.

4.3 Methods

4.3.1 Genetic toggle switch

The basic genetic toggle switch consists of two mutually repressing genes as shown in Fig. 4.1 along with an additional system to toggle the states. As shown in previous studies, with the right combination of parameters, the toggle switch will stay in one of two stable states, each characterized by a high concentration of one of the repressor proteins, and strong repression of the other. The toggle switch can now be induced to switch states using two possible strategies for inducing a transition: decrease the level of highly expressed protein [5], [36] or increase the expression of one of the repressed proteins (Fig. 4.1) using an additional inducible system [36]. For a model which utilizes the latter protocol we obtain a system of four differential equations [36] after including a load. The load may be a protein, a small molecule or a binding site on DNA such that the bound complex prevents the repressor from binding to and repressing its conjugate promoter. In order to make the simplest and the most general model, we have assumed here that the repressors reversibly bind the load only in one copy. We assume that the total load L1T is a constant, L1 is the free load and conservation gives us the bound load as L_{1T}–L_1.

$$\frac{du}{d\tau} = \alpha_1 + \frac{\beta_1{}'}{1+v^n} - u - k_{on1}{}'[L_{1T}]u\ell_1 + k_{off1}{}'[L_{1T}](1-\ell_1) \tag{1}$$

$$\frac{dv}{d\tau} = \alpha_2 + \frac{\beta_2{}'}{1+u^n} - v - k_{on2}{}'[L_{2T}]v\ell_2 + k_{off2}{}'[L_{2T}](1-\ell_2) \tag{2}$$

$$\frac{d\ell_1}{d\tau} = -k_{on1}{}'k_1u\ell_1 + k_{off1}{}'k_1(1-\ell_1) \tag{3}$$

$$\frac{d\ell_2}{d\tau} = -k_{on2}{}'k_2v\ell_2 + k_{off2}{}'k_2(1-\ell_2) \tag{4}$$

These four equations are presented in de-dimensionalized form, with u, v, l_1, l_2 representing the dimensionless concentrations of Repressor 1, Repressor 2, Load1 and Load2 respectively and τ the de-dimensionalized time. The basal parameter values that we use are as follows: $\alpha_1=\alpha_2=0.2$; $\beta_1'=\beta_2'=4$; $n=3$; $k_{on1}'=k_{on2}'=0.5$; $k_{off1}'=k_{off2}'=0.5$; $k_1=k_2=1$; $[L_{1T}]$ and $[L_{2T}]$ are variable. Note that Equations (1) and (2) without the last two terms incorporating the load are the standard equations for analyzing the toggle switch that have been widely used in both empirical and theoretical work [5], [36]. These equations are discussed in more detail in Supplementary Text S1 Section 1.1. The derivation of this model follows that of Kobayashi et al [33]. All parameters excluding load binding rates were sourced from Kobayashi et al [36]; extensive parameter sensitivity of the load binding rates was performed and are discussed in Appendix II: Supplementary Text S4.1 section 1.4 and Figs. S4.1, S4.2, Table S4.1 and Figs. S4.15 and S4.16. The effect of a load arises from the binding competition between the promoter where the repressor binds and the load. This competition is not directly incorporated into the Hill function, since the binding step with the promoter is not explicitly modeled and is treated in an effective way. In reality however the concentration of the promoter is so small compared to that of the load, that the use of Hill functions is justifiable [37]. There are possibly exceptional cases such as a high copy number of plasmids compared to load concentrations where this assumption does not apply. Note that the Hill function is an effective phenomenological equation describing gene transcription and protein production, and standard Law of Mass Action (LMA) methods to derive the Hill functional form may not apply for many transcription factors that nevertheless show Hill kinetics [38]. Thus it is preferable to use Hill function forms for this analysis.

To calculate transition times, we first start the system in one state, say high Repressor 1. After the system has reached steady-state, we add a constant concentration of the inducer and measure the time taken for Repressor 2 to go from 10% of its maximum value to 90% of its maximum value. This is the “rise time”. Similarly the “decay time” is the time taken for Repressor 1 to go from 90% of its maximum value to 10% of its maximum value. The level of the inducer remains fixed.

In practice the inducer may decay and the transition would depend upon there being inducer present for a sufficiently long time to induce transition. In such cases the amount of inducer required may be of interest. When the inducer is applied as a bolus with a first order decay rate, it appears as an exponentially decaying pulse. We thus included a fifth differential equation governing the amount of Inducer.

$$\frac{dI_1}{d\tau} = -d_I I_1 \tag{5}$$

Here d_I is the ratio of the inducer degradation constant to the repressor degradation constant. We used Eq. 5 only when estimating the amount of inducer required to switch states for different loads and different decay rates of the load (Appendix II: Supplementary Text S4.1 section 1.4 and Supplementary Tables S4.3, S4.4).

A genetic toggle switch can be induced to change states by the alternative method of repressing the highly expressed repressor, and in fact the original toggle switch used this form of induction [5], [33]. We repeated our calculations for the basic model for the case of alternative induction, but found no qualitative differences. The alternative induction model along with the equations is detailed in the Supplementary Text S4.1 section 1.4.

Equations 1–4 assume that the load itself stays in steady state during the switching of the toggle between one state and another. However in reality if the load is another protein, it is also synthetized and degraded by the cell, and therefore its level could be dynamic. We also simulated this situation by incorporating a synthesis and a degradation rate for each load. This resulted in Equations 3 and 4 being replaced by:

$$\frac{d\ell_1}{d\tau} = -k_{on1}{}' k_1 \frac{k_{b2}}{k_{d2}} u\ell_1 + k_{off1}{}' k_1 \frac{k_{b1}}{k_{d1}} (c_1) + \frac{k_{d1}}{\delta} - \frac{k_{d1}}{\delta} \ell_1 \tag{6}$$

$$\frac{d\ell_2}{d\tau} = -k_{on2}{}' k_2 \frac{k_{b2}}{k_{d2}} v\ell_2 + k_{off2}{}' k_2 \frac{k_{b2}}{k_{d2}} (c_2) + \frac{k_{d2}}{\delta} - \frac{k_{d2}}{\delta} \ell_2 \tag{7}$$

Here c_1 is the load-repressor complex and k_{b1} and k_{d1} are the synthesis and degradation rates respectively for Repressor 1, and correspondingly for Repressor 2. The parameters are defined in Appendix II: Supplementary Text S4.1, section 1.5. Since the total load is no longer conserved, we need to include additional equations for the load repressor complex.

$$\frac{dc_1}{d\tau} = k_{on1}{}' k_1 u\ell_1 - k_{off1}{}' k_1 c_1 \tag{8}$$

$$\frac{dc_2}{d\tau} = k_{on2}{}' k_2 u\ell_2 - k_{off2}{}' k_2 c_2 \tag{9}$$

Our model assumes that when the repressor protein is bound to the load, it is protected from degradation. However it is also possible that even when the protein is bound to the load, it can still degrade. To check the impact of removing the protection assumption, we also consider an additional model where the repressor can still degrade with the same rate constant when bound to the load. The equations for that model are slightly modified versions of the equation above, and are presented in detail in Appendix II: Supplementary Text S1, section 3.2.

We conducted parameter sensitivity analysis on models utilizing both forms of induction; these did not show any qualitative change on wide variation of key parameters (Appendix II: Tables S4.1, S4.2, S4.3, S4.4 and Supplementary Text S4.1).

4.3.2 Toggle with positive feedback

A positive feedback was added to the R1 side of the toggle switch as an inducible promoter with a Hill coefficient of 1. We assumed that the positive feedback acted on the same promoter as the repression, resulting in a composite term for production of R1 from promoter 1 where ρ is the strength of positive feedback.

$$\frac{dR_1}{dt} = \alpha_1 + \frac{\rho R_1 + \beta_1}{1 + R_1/k_5 + R_2^{n_2}/k_2^{n_2}} - d_1 R_1 \tag{10}$$

The derivation of this equation can be found in Appendix II: Supplementary Text S4.1, section 1.6.1. As before, α1 is the leaky production of R1 while α1+β1 represents the activity of the promoter in the absence of repression or positive feedback. We chose k2 and k5=1, d1=0.2, and for the figures in the main paper we chose ρ=3.5. We address other values of the positive feedback in Appendix II: Fig. S4.6 and the Supplementary Text S1, section 1.6.2.

4.3.3 Stochastic simulations

We perform stochastic simulations and histogram the concentrations of the repressor proteins to construct their probability distribution. The quasi-potential of the toggle is given by the negative logarithm of this probability distribution [39]. In order to construct the probability distribution we make use of the phenomenon of noise-induced switching. Recent theoretical work has shown that multiplicative noises due to stochastic fluctuations can induce switching

[40]–[42]. Experimental results demonstrate bimodal populations that correspond with theoretic predictions arising from noise-induced switching [41].

Stochastic simulations were carried out using a modified Gillespie algorithm using the standard rate expressions for every reaction (Table S4.5). We chose a reaction volume that would correspond to a small number of molecules in the system. Stochastic fluctuations then drive the system to transition between states rapidly, allowing us to collect sufficient data points. In order to make sure that the system was not being biased by the small volume, we also repeated the calculations for a five times larger volume (and hence molecule number) and found qualitatively similar results (Appendix II: Fig. S4.4).

For the positive feedback toggle switch the same equations were used except for the repressible production of Repressor 1, where we used instead the rate expression given by the right hand side of Eq. 10 in the Monte Carlo simulations.

4.3.4 Ras-kinase system

For our study we adapted the minimal model of the Ras switch proposed by Das et. al. [30] with the addition of a reversibly binding load in the form of the Raf protein (Fig. 1C). The model contains three proteins, Ras, which exists as RasGDP or RasGTP, SOS, the guanine exchange factor (GEF) that catalyzes the transformation from RasGDP to RasGTP and a GTPase, RasGAP. SOS on its own has very low GEF activities. However, the activity of the GEF pocket is strongly influenced by the binding state of an allosteric pocket in Cdc25 domain [29], [30]. When the allosteric pocket is bound by RasGDP, the GEF activity is increased by 5 times. If the allosteric pocket is bound by RasGTP, its GEF activity is increased by 75 times. In this way, RasGTP can upregulate its own production rate by binding to SOS, thus constituting a positive feedback loop. RasGTP is deactivated by GTPase's such as RasGAPs that are constitutively present.

After Raf binds RasGTP, the complex catalyzes the phosphorylation of Raf leading to a phosphorylation cascade. For this study we ignore Raf activation and only consider the effects of Raf as a binding partner for RasGTP. The Das paper [30] also models the systems using Michaelis-Menten (MM) forms for the actions of the enzymes which is quite standard for modeling systems of enzymatic reactions. However since in this model the load competes not with a promoter, as in the toggle switch, but with another protein, it is possible that the quasi-steady state assumption of the MM form could be introducing some inaccuracies in the results. To account for this possibility we wrote the entire model using the Law of Mass Action. We separately simulated the model using the MM functional forms (Supplementary Text S1 section 2 and Figs. S7 and S9). The equations for the MM forms are listed and discussed in detail in the Supplementary Text S1. The reactions and rate constants for this model are listed in Table S6 and Table S7.

We use the following notations for the species involved in the system:

$$\begin{aligned}
&x_1 \equiv [SOScat];\ x_2 \equiv [RasGDP];\ x_3 \equiv [RasGTP];\\
&x_4 \equiv [SOScat(RasGDP)];\ x_5 \equiv [SOScat(RasGTP)];\\
&x_6 \equiv [SOScat(RasGDP) : RasGDP];\\
&x_7 \equiv [SOScat(RasGTP) : RasGDP];\ x_8 \equiv [RasGAP];\\
&x_9 \equiv [RasGAP : RasGTP];\ x_{10} \equiv [Raf];\ x_{11} \equiv [RasGTP : Raf]
\end{aligned}$$

$$\frac{dx_1}{dt} = -k_{on1}x_1x_2 + k_{off1}x_4 - k_{on2}x_1x_3 + k_{off2}x_5 \tag{11}$$

$$\begin{aligned}\frac{dx_2}{dt} = &-k_{on1}x_1x_2 + k_{off1}x_4 - k_{on3}x_2x_5 + k_{off3}x_7\\ &-k_{on4}x_2x_4 + k_{off4}x_6 + k_{cat5}x_9\end{aligned} \tag{12}$$

$$\begin{aligned}\frac{dx_3}{dt} = &-k_{on2}x_1x_3 + k_{off2}x_5 + k_{cat3}x_7 + k_{cat4}x_6 \\ &-k_{on5}x_3x_8 + k_{off5}x_9 - k_{on6}x_3x_{10} + k_{off6}x_{11}\end{aligned} \tag{13}$$

$$\frac{dx_4}{dt} = k_{on1}x_1x_2 - k_{off1}x_4 - k_{on4}x_2x_4 + k_{off4}x_6 + k_{cat4}x_6 \tag{14}$$

$$\frac{dx_5}{dt} = k_{on2}x_1x_3 - k_{off2}x_5 - k_{on3}x_2x_5 + k_{off3}x_7 + k_{cat3}x_7 \tag{15}$$

$$\frac{dx_6}{dt} = k_{on4}x_2x_4 - k_{off4}x_6 - k_{cat4}x_6 \tag{16}$$

$$\frac{dx_7}{dt} = k_{on3}x_2x_5 - k_{off3}x_7 - k_{cat3}x_7 \tag{17}$$

$$\frac{dx_8}{dt} = -k_{on5}x_3x_8 + k_{off5}x_9 + k_{cat5}x_9 \tag{18}$$

$$\frac{dx_9}{dt} = k_{on5}x_3x_8 - k_{off5}x_9 - k_{cat5}x_9 \tag{19}$$

$$\frac{dx_{10}}{dt} = -k_{on6}x_3x_{10} + k_{off6}x_{11} \tag{20}$$

$$\frac{dx_{11}}{dt} = k_{on6}x_3x_{10} - k_{off6}x_{11} \tag{21}$$

Moreover, four of the basic protein species along with the complexes they participate in have associated conservation laws. These are as follows:

$$SOS_T = x_1 + x_4 + x_5 + x_6 + x_7 \tag{22}$$

$$Ras_T = x_2 + x_3 + x_4 + x_5 + 2x_6 + 2x_7 + x_9 + x_{11} \tag{23}$$

$$GAP_T = x_8 + x_9 \tag{24}$$

$$Raf_T = x_{10} + x_{11} \tag{25}$$

In the Ras model too we implicitly assume that when RasGTP is bound to Raf, it is protected from de-activation by a RasGAP. We also study the effects of relaxing this assumption on both the LMA and the PSSA models. The modifications to the original model are detailed in Appendix II: Supplementary Text S1 section 3.3.1.

We used XPPaut to perform a bifurcation analysis of the Ras switch with changing levels of SOS, with and without a load. The quasi-potential landscape does not provide useful insights into load induced modulation of the Ras switch and hence the probability distributions are not reported.

4.4 Results

4.4.1 The bistability properties of the toggle switch do not change unless the repressor can degrade when bound to the load

The presence of a binding partner for either Repressor 1 or Repressor 2 (which we refer to thereafter as the load) introduces new terms in the differential equations describing the toggle switch, i.e. the last two terms in Eq. 1 and in Eq. 2, as well as two new equations, Eq. 3 and 4, in the dynamical system. However it can be easily seen that in steady state Eq. 3 and 4 are also independently set to zero, and therefore do not affect the bifurcation properties of the switch. Even in the case of a dynamic load, since Eq. S13 and S14 are set to zero to ensure the load-repressor complex is in steady state, the additional terms in Eq. S9 and S10 are also zero. Thus the load makes no difference to either the bistability of the switch or to the parameter values where the bistability is seen.

The exception is when the repressor molecule can degrade even when bound to the load, which may be relevant in some experimental situations. As Fig. 4.2A shows, when a load is added symmetrically to both sides of the toggle switch, the two stable states approach each other and eventually annihilate, leaving a monostable system. Fig. 4.2B shows that when a load is added only to one side, the system again goes from bistable to monostable at some critical value of the load. In effect, the upper stable point vanishes and is no longer accessible due to leakage of the repressor affected by the load.

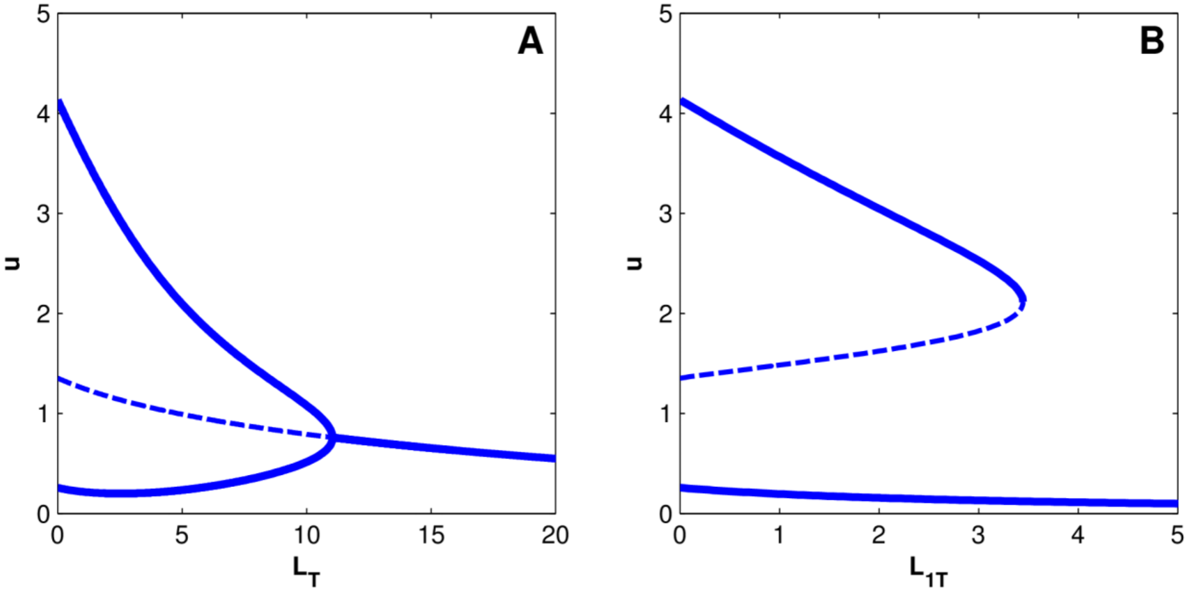


Figure 4.2. Bifurcation diagram of the genetic toggle switch when the repressor can decay from the load-repressor complex.
The thick lines are stable steady states, the dashed lines are unstable steady states. (A). A load is added symmetrically to both sides of the toggle. The stable states of only one Repressor molecule with respect to the load are shown. With zero load the toggle switch is bistable with well separated steady states. As the load increases, the two stable states approach each other and the unstable state, and eventually merge in a bifurcation at a critical value of the load. The system is monostable beyond this critical value. (B) A load is added only to Repressor 1. The high state of Repressor 1 approaches the unstable steady state as the load increases and merges with it at a critical value of the load, leaving only the lower state accessible to the system.

The reason for the change in steady state behavior is made clear on examining the equations of the system. Here we need to incorporate additional reactions that represent the decay of the repressor-load complex into the load alone. This leads to an additional term in the equation for the load and the repressor-load complex (Eq. S44 and S45). However this term does not appear in the equation for the repressors, which continue to be governed by Eq. 1 and Eq. 2. As a consequence in the steady state, the additional terms in Eq. 1 and 2 no longer equal zero and the steady state properties of the switch are influenced by the presence of the load.

As can be seen from an examination of the chemical reaction system, this mechanism of abrogation of bistability arises whenever the load-repressor complex participates in a non-reversible (from the repressor's point of view) chemical process that leads to an unbalanced leakage of the repressor from its function as a repressor by the presence of the load. A more interesting example of such a process could be provided by a chemical reaction system where the load is an enzyme for one of the repressor molecules, which is transformed by the enzymatic action into a protein no longer capable of repression. The mathematical analysis of this case is exactly the same as the model we are currently discussing hence we do not consider it separately here.

However a load can significantly change the dynamic response of the basic genetic toggle switch as we shall see below. We examined two different measures of dynamic response, response time for state switching and the amount of inducer required for state switching.

4.4.2 The response time for state switching of the toggle switch increases

We measured two response times, the rise time which quantifies the time taken for the concentration of Repressor 2 to increase from its low or zero level in state 1 to its high level in state 2, and the decay time which measures the time taken for Repressor 1 to decay from its high level in state 1 to its low level in state 2, in both cases in response to a constant inducer. Specifically the rise time measured the time to go from 10% to 90% of the steady state maximum, while the decay time measured the time to go from 90% to 10% of the steady state maximum. These measurements were made using the deterministic model in the cases when the load was applied only to one side and to both sides of the switch.

We found that both the rise time and the decay time increase with increasing load concentration. Interestingly, this relationship was approximately linear in all cases (Fig. 4.3A & B). The slope of the linear relationship represents the increase in response time due to unit increase in load. We found that the slope of the line was larger when the load was applied to the opposite side of the system before the switching rather than the same side (Fig. 4.3A), indicating that it is harder to switch out of a state without a load to a state with a load than the reverse. However when a load was applied to both sides, the slope of the linear fit was higher than when the load was only on the opposite side, suggesting that both the "opposite side" and the "same side" delays are operating.

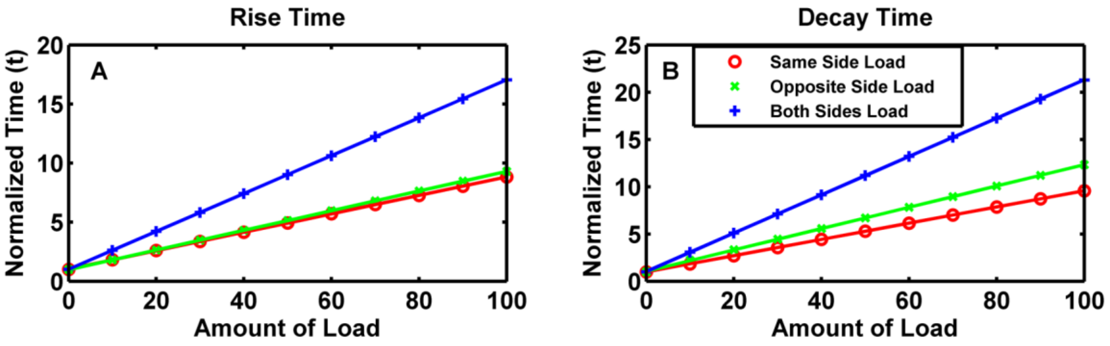


Figure 4.3. Effects of a load on transition times of the basic toggle switch.
(A). The time taken to reach 90% of maximum value for the protein undergoing a low-to-high transition as a function of the Load, normalized by the steady-state amount of Repressor 1. Normalized time is a unit-less number defined by the transition time (rise or decay) of the system at a given loading condition divided by the transition time (rise or decay) of an unloaded system. (B). The time taken for the concentration of the protein undergoing a high-to-low transition to reach 10% of its maximum value. The x- and y- axes are the same as for the previous panel.

While we also found an approximately linear relationship between the decay time and the concentration of the load, there was little difference between the decay times for the state with the load ("same side load") and the state without a load ("opposite side load") at our base parameter values. Thus the load affects rise time and decay time differently. When a load was applied to both sides of the switch, the slope of the decay time linear fit was larger, again indicating the operation of both delays.

We tested these results by changing parameter values for the binding of the load (Appendix II: Table S4.1) and found that in all cases we obtain a good linear fit for the response time. For the rise time, the slope was uniformly larger when the load was applied to the opposite side as compared to the same side, and it was the largest when loads were applied on both sides. For the decay time, the slope could be larger or smaller when the load was applied to the opposite side of the decaying state compared with the same side, but it was always larger than both when a load was applied on both sides. The slope depended non-monotonically upon the dissociation constant (Kd) of the binding between the repressor protein and the load, with both low Kd and high Kd having a smaller effect that those in between (Appendix II: Fig. S4.1). This was because when the Kd was low, i.e. strong binding, the concentration of the load-repressor complex was unaffected by the state of the switch. However when the Kd was high, the maximum concentration of the load-repressor complex was smaller, thereby having a lesser effect on the system (Appendix II: Fig. S4.2). Thus response times are maximized when the load acts as a dynamic sink, i.e. it takes up newly synthesized repressor when the state changes from the unloaded to the loaded side, and releases the bound repressor when switching from the loaded side to the unloaded side.

Previous studies of response times of biochemical networks with and without a load have also seen monotonic increases in the response time of simple transcriptional circuits [37]. However the extremely consistent approximately linear response we see under wide variation in parameter values is extremely intriguing.

An increase in response time should also imply that the concentration of inducer required to shift states should also be affected, especially when it can decay. In accordance with this expectation we also found that the concentration of inducer required to switch states increased exponentially with increasing load, as seen in Appendix II: Table S4.2. The parameter of the exponential fit was dependent on the inducer decay rate, indicating that the amount of time the inducer remains above a threshold is the key factor governing the switching. We find that this response to a load is unaffected by the mode of switching the toggle, and induction by repression of the current state yields the same qualitative results (Appendix II: Table S4.2 & S4.3).

In our analysis so far we have assumed that the total concentration of the load is fixed. We now analyze the case when the load is generated by a constitutively active promoter and can decay at a first order rate. We find that in this case too the qualitative features of the transition time remain the same as the toggle switch with a fixed load, i.e. it was approximately linear in all parameter regimes tested (Appendix II: Supplementary Text S4.1 section 1.5, Fig. S4.3 and Table S4.4). The reason why we do not see a difference from the basic toggle switch is that the transition times ultimately measures time between steady states, and we wait for the system to come quite close to the steady state value (90%). Thus the concentration of the load has also reached a steady state value and the system behaves as it would with a fixed load.

We also tested the response times when the repressor can leak away from the systems after binding with the load. Here we find that (Fig. 4.4) when a load is applied to the same side, the rise time continues to increase monotonically linearly with the load but the decay times decreases monotonically with the load. However when a load is applied to both sides, we find a negative linear relation between the transition times for both rise and decay and the load.

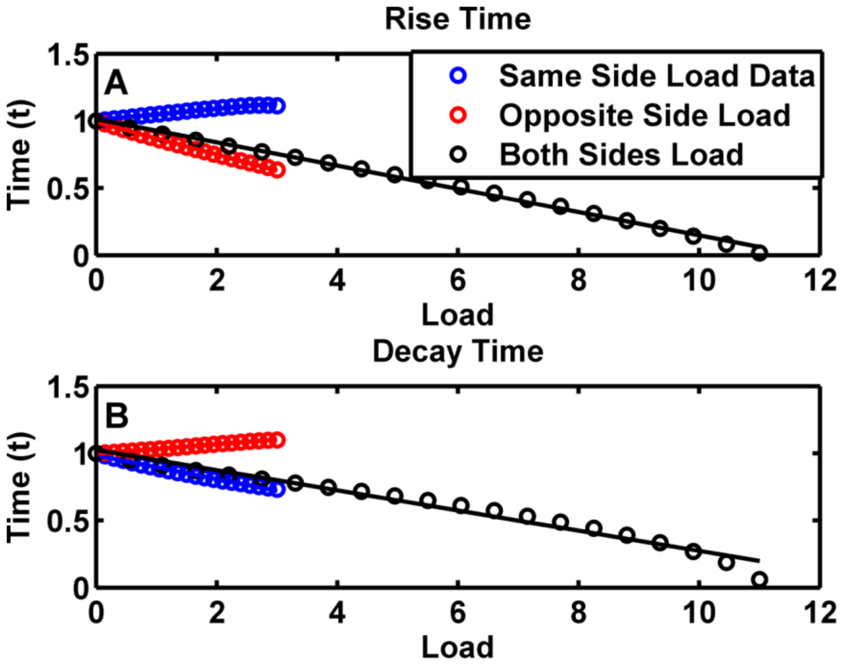


Figure 4.4. Effects of a load on transition times of a toggle switch without the protection assumption.
(A). The time taken to reach 90% of maximum value for the protein undergoing a low-to-high transition as a function of the load. The system is de-dimensionalized as described in Supplementary Text S1 section 1.1 and 3.2.1. (B). The time taken for the concentration of the protein undergoing a high-to-low transition to reach 10% of its maximum value. Note that the linear relationship for both-sided load transition times, and same-sided decay time, and opposite-sided rise time has a negative slope. The relationship for same-sided rise time and opposite sided decay time has a very small, but positive slope.

The reasons for the change in behavior is because as we saw previously, when the repressor can leak away from the repressor-load complex, a load has a dramatic effect on the bistability properties of the switch, abrogating bistability very quickly (Fig. 4.1). When only one repressor has a load, the high state of that repressor approaches the unstable state, indicating a decrease in the domain of attraction. Shifting out of that state thus becomes easier with increasing load. When both sides have loads, both stable states approach the unstable state, therefore shifting out of either state becomes easier, and both transition times decrease.

4.4.3 Dramatic changes in the potential energy landscape and probability distributions of the toggle switch

The modulation in the dynamic properties of the basic genetic toggle switch discussed above suggests that the load has altered the potential energy landscape of the toggle switch, making it harder to switch. For two-dimensional and higher systems, such as the toggle switch, analytical methods to construct the potential landscape are not available, but a quasi-potential can be constructed from the probability distribution function of the concentrations of the repressor molecules, where the quasi-potential is given by the negative of the natural logarithm of the probability distribution [43], [44]. To calculate this we performed Monte Carlo simulations of the toggle switch using a Gillespie type algorithm elaborated in the Methods section. When the toggle switch is symmetrically balanced, both the probability distribution function and the potential energy landscape are completely symmetric. If the system is started in State 1, random fluctuations can drive it into State 2 and vice versa. The probability distribution can then be constructed by counting the frequencies of these random fluctuations. However since the genetic toggle switch can be very stable, a numerical computation of the potential energy landscape requires impractically long simulation times (as we show below). While computational methods to sample rare trajectories in such cases exist, they are very sensitive to choices of parameters [42], [45]. We developed a computational protocol in order to numerically obtain the probability distribution function of both protein concentrations and the transition times. We chose an appropriate volume for the genetic toggle switch such that exactly the same parameters as in the deterministic simulations led to the operation of the toggle switch with only a small number of proteins. The toggle remains bistable in this regime but the small protein numbers vastly increases spontaneous stochastic fluctuations arising out of multiplicative noise in the system and allows the simulation to explore parameter space and collect enough data.

Our simulations showed that the switch switched states a large number of times. In order to account for differences in the time step in different states, the probability density function of the concentrations was constructed using a time trace collected after approximately 1 second intervals. As Fig. 4.5 shows, for a symmetric switch we obtain a symmetric bi-modal probability distribution that corresponds to a double-well potential.

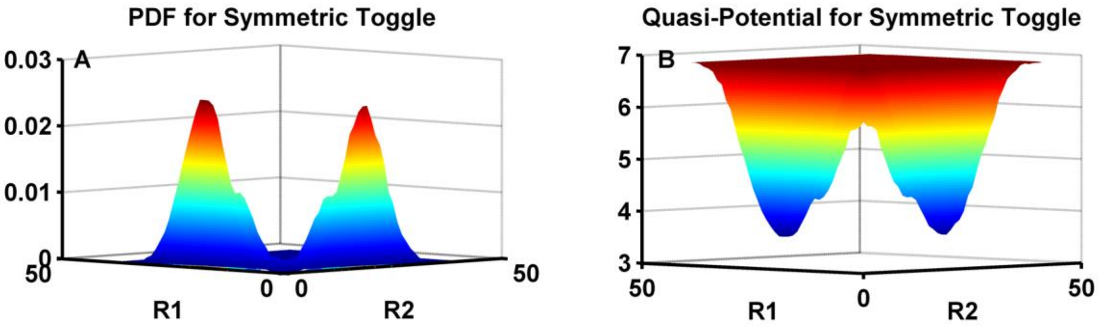


Figure 4.5. The probability distribution function and the quasi-potential of the genetic toggle switch without a load.
(A). The probability distribution function of a toggle switch without a load. The x- and y- axes here represent the number of molecules of Repressor 1 and Repressor 2 respectively, while the z-axis is the frequency of its occurrence. Note that the distribution is symmetric as expected. (B). The quasi-potential of the symmetric toggle switch, showing the symmetric double-well potential constructed by taking the negative logarithm of the probabilities shown in (A). A small offset of 0.001 was added to the probabilities to prevent taking the logarithm of zero. This does not change the shape of the well.

When we add a load to the system asymmetrically, in the form of a binding partner for the Repressor 1, we find that the probability distribution becomes extremely skewed, and the total weight of the probability distribution corresponding to the other side, i.e. Repressor 2, dramatically increases (Fig. 4.6A). This indicates that the underlying double well potential has become skewed and the state 2, corresponding to high Repressor 2, has increased its stability at the cost of State 1 (Fig. 4.6C). When a load is applied to both sides symmetrically, the concentration probability distribution reverts to a symmetric bimodal distribution corresponding to a symmetric double-well potential (Fig. 4.6B & D).

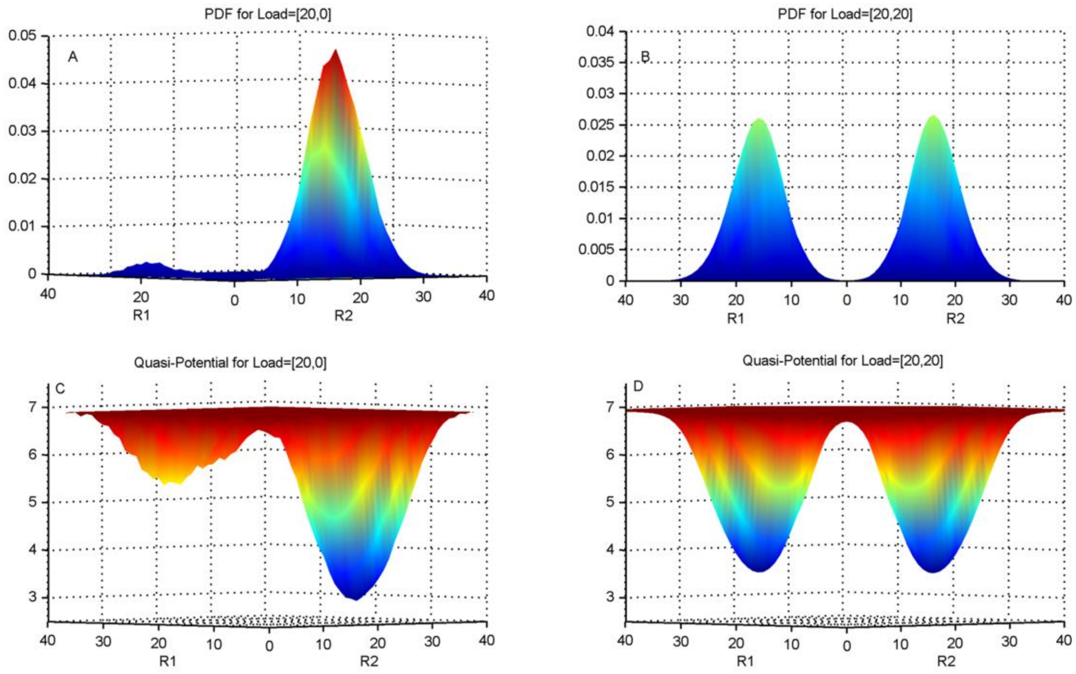


Figure 4.6. The probability distribution function and quasi-potential of a toggle switch with a load.

The 3-dimensional plot is viewed with the xy-plane horizontal for better contrast. The x- and y-axes are numbers of molecules of R1 and R2 while the z-axis is either probabilities or the quasi-potential. (A). The probability distribution function (pdf) of the toggle switch of Fig. 5 but now with a load of 20 molecules on Repressor 1 (R1). (B). The pdf of the toggle switch with a load of 20 molecules on R1 and 20 molecules on R2. (C). The quasi-potential of the toggle with a load of 20 molecules on R1, i.e. corresponding to panel A. (D). The quasi-potential of the toggle with equal loads of 20 molecules on each repressor, i.e. corresponding to panel B.

In order to test this directly we calculated the distribution of lifetimes in state 1 and the lifetimes in state 2. As shown in Fig. 4.7, when the switch is symmetric with no load, the lifetime distribution is exponential, as should be expected for a simple two-state system. However when the load is applied to Repressor 1, the probability distribution of the lifetime in state 2 increases dramatically. The average lifetime of state 1 also increases but only by a very small amount. The time spent in state 2 does not appear to saturate, and continues to increase with increasing load. When loads are applied symmetrically to both sides, the lifetime histogram in Fig. 4.7 indicates that both sides have been stabilized since the system spends significantly longer time in each state. Note that in an equilibrium system this would have been indicated by the deepening of the potential well. However in non-equilibrium systems the potential well picture does not completely capture the dynamics and there is an additional contribution from a “curl flux” [43], [46] that needs to be taken into account. For our purposes calculating both the distribution of concentrations and the distributions of lifetimes captures the dynamics of the toggle switch.

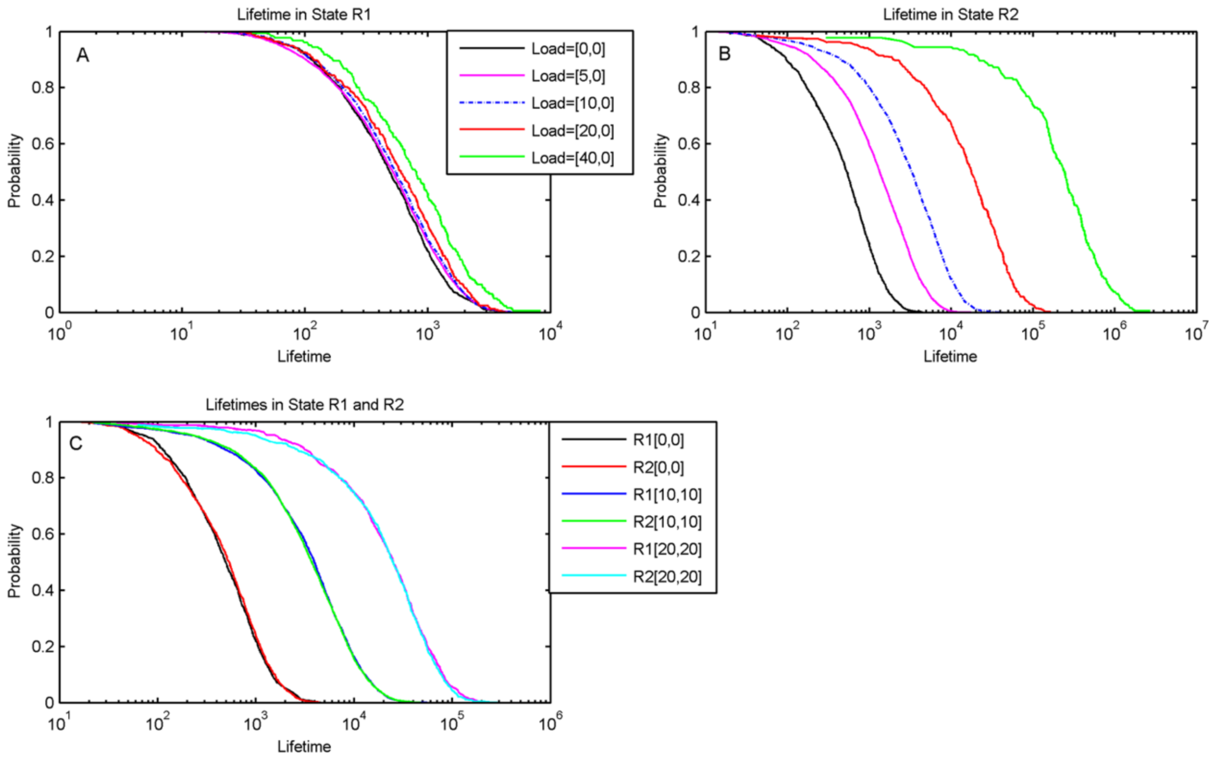


Figure 4.7. Distribution of the lifetimes of the toggle switch with and without loads. The time the system spent in either state R1 or state R2 was calculated from the time trace of the stochastic simulations and a histogram made of the results. The histogram is shown on a semi-log plot to accommodate the data on a single chart. (A). Lifetimes in State R1. The unloaded state is the solid curve that is to the extreme left of the others, showing that the lifetimes in state R1 increase slightly on addition of load on R1 alone due to the "same side effect". (B) Lifetimes in State R2 when load is on R1. The solid curve on the extreme left is the unloaded state. There is a significant increase in lifetimes due to the "opposite-side effect" of the load on R1. (C). Lifetimes with a balanced load, showing that both the states R1 and R2 get stabilized with a significant increase in lifetimes on addition of a small load on both sides. Note that the distributions for R1 and R2 for equivalent cases coincide as should be expected.

To test whether our results change for higher protein concentrations, we increased protein concentrations about fivefold and recalculated the probability distribution function. We find that our qualitative results remain robust despite the increase in protein concentrations (Appendix II: Supplementary Text S4.1 section 1.3 and Fig. S4.4). Switching between states is rare at these protein numbers, with a mean residence time in state R1 for the unloaded switch being approximately 6×10^5 min against about 700 min for the basal case considered, a difference of almost three orders of magnitude. However as for the basal case, the quasi-potential landscape skews significantly with the addition of a load on the switch.

4.4.4 "Opposite Side effect" dominates the load effect in the basic toggle switch

These results allow us to interpret the dynamic results that we obtained earlier. If the system is in state 2 and there is a load on state 1, a transition requires an increase in Repressor 1 concentration in order to suppress the production of Repressor 2. A load on Repressor 1 however competes with the promoter of Repressor 2 for binding with Repressor 1, and thereby reduces the effective concentration of Repressor 1. This effectively stabilizes state 2. The dynamic analysis shows that state 1 not only remains an attractor state but in fact it takes a longer time, and more inducer, to shift out of state 1 as compared with the no-load situation. This is because the load also acts as a reservoir for Repressor 1, and in fact increases its total concentration. This slows down the transition to state 2. Interestingly this "same side effect" is generally weaker than the "opposite side effect" above. In agreement with this picture, the stochastic simulations show that the distributions of lifetimes in state 1 broaden slightly on addition of a load.

If the load is present symmetrically on both sides, the concentration histograms in Fig. 4.6 and the time histograms in Fig. 4.7 indicate that both states have been stabilized, due to a combination of the 'same side' and the 'opposite side' effect now acting together to stabilize each state of the switch. In the dynamical simulations this is seen by the increased slope of the response time line for the case of a load on both sides. Results for additional parameter values are shown in Appendix II: Fig S4.15 and Fig S4.16.

4.4.5 Positive feedback moiety makes toggle switch tunable

When a positive feedback moiety is introduced in the toggle switch, we again see a linear relationship between the rise time and the decay time of the two states of the switch and the load (Appendix II: Fig. S4.5). Therefore here too the load appears to be skewing the underlying potential landscape of the switch. Using stochastic simulations we constructed the probability distribution function of this toggle switch as described above. We found that even in the absence of a load, when a positive feedback moiety is introduced on one side of a toggle switch, the probability distribution for the toggle switch, and hence the quasi-potential landscape, becomes extremely skewed in favor of the state with positive feedback as shown in Fig. 4.8A. Even with no load on the system, the switch is biased to State 1 and the lifetime spent in State 1 is much longer than in State 2. If a load is added to R2, the opposite side effect additionally favors State 1. If a load is added to R1 however, the opposite side effect favors State 2 (Fig. 4.8B). It is possible to balance these effects resulting in a more even distribution by adjusting the load on R1 and the strength of positive feedback. As the load on R1 is increased beyond this balance point, the opposite side effect dominates and the probability distribution becomes skewed toward State 2 (Fig. 4.8C). As the opposite side effect increases with increasing load, the lifetime in State 2 also increases in agreement with the findings for the regular toggle switch (Fig. 4.8D). The lifetime in State 1 also increases by a smaller amount, as for the regular toggle switch (Fig. 4.8E).

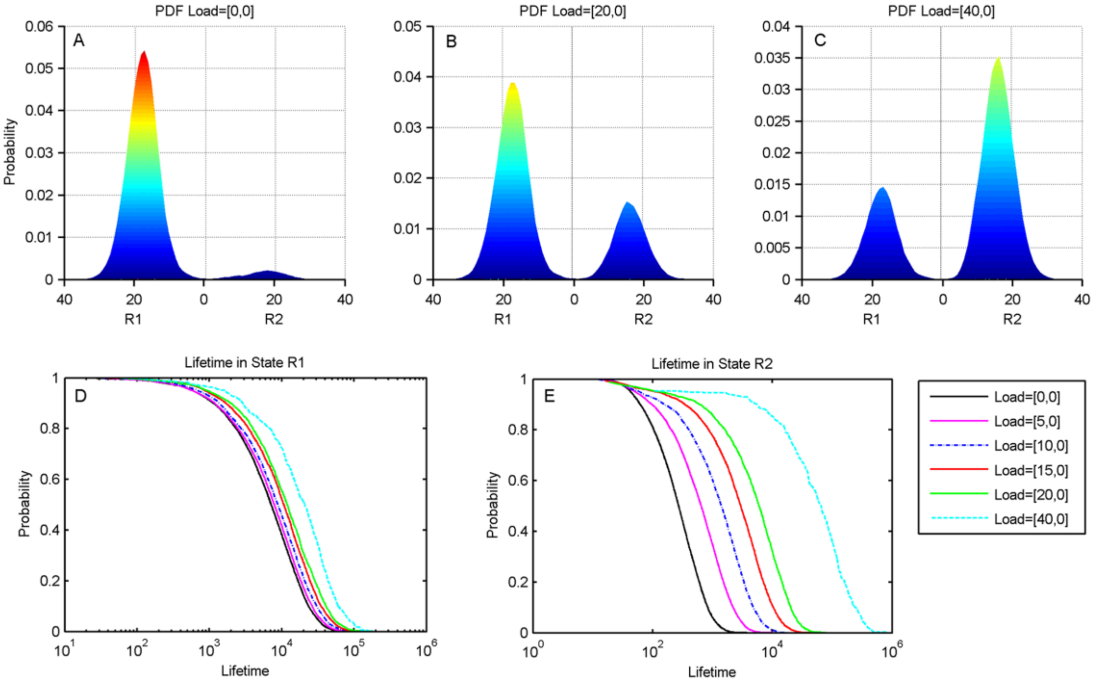


Figure 4.8. The genetic toggle switch with a positive feedback motif on Repressor 1 (R1). (A). The probability distribution function (pdf) with no load. The positive feedback on Repressor 1 leads to a pdf skewed in favor of R1. (B). The pdf with a load of 20 molecules on R1 showing the increase in the weight of R2 due to the "opposite side effect". (C). The pdf with a load of 40 molecules on R1. This load is more than enough to skew the pdf in favor of state R2. (D). Histogram of lifetimes in R1 with varying levels of load on R1. Comparison with panel A shows that the unloaded state has been stabilized by the positive feedback. Note that the lifetimes increase very slightly due to the "same side effect". (E). Histogram of lifetimes in R2 with varying levels of load on R1. The unloaded case is the curve on the extreme left. Note the initial asymmetry in the lifetime distribution due to the positive feedback, as well as the large increase in lifetimes with the inclusion of a load.

For the toggle switch with the positive feedback moiety, we can also check the consequences of allowing repressor leakage through the repressor-load complex. As shown in Appendix II: Fig. S4.13, this addition to the system affects the steady state properties of the switch and bistability is abrogated after the load increases beyond a critical value, when load is present for both sides or only one side.

4.4.6 Loads fundamentally transform positive feedback based switches in signal transduction

The RasGTP system shows a bistable transition from a low RasGTP state to a high RasGTP state as the activating signal, in our case the number of SOS molecules, are varied. As Fig. 9 shows, a system with no Raf shows a classic Z-shaped bifurcation diagram with two bifurcations as SOS is varied. The first bifurcation marks the transition from a monostable low-RasGTP state to a bistable system with a “high” RasGTP state (and an unstable intermediate state). The second bifurcation marks the transition from the bistable state to another monostable state with a high concentration of RasGTP.

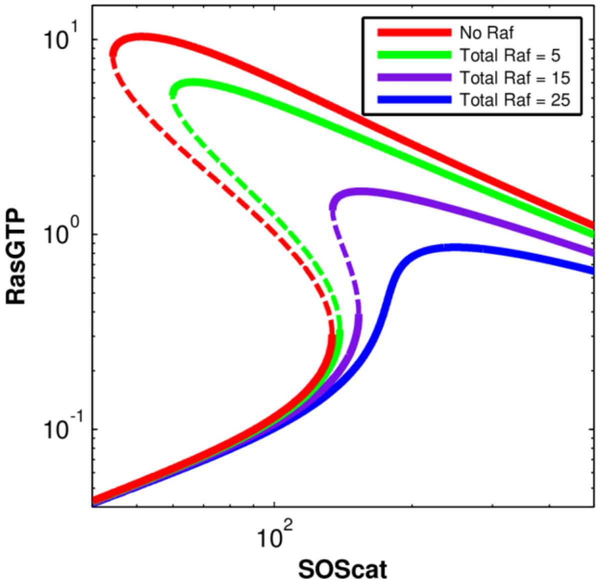


Figure 4.9. Bifurcation diagram of the Ras switch with different levels of Raf (load) on the system.
The total number of SOS in the simulation box is used as the parameter being tuned, which varies from 0 to 1000. For Raf=0, Raf=10 and Raf=30, there are two bifurcations as SOS is increased. In the first bifurcation a new high valued stable steady state appears along with the low valued stable steady state. In the second bifurcation, the low valued stable state disappears leaving behind only the high valued state. The dotted line marks the unstable steady state that also comes into existence in the bistable region. As total Raf increases, the two bifurcations approach each other. When Raf=50, the system has lost both of its bifurcations and is characterized by a single stable steady state at all values of Raf.

When Raf is added to the system, the bifurcation diagram changes and the two bifurcations start approaching each other. This is because the effect of adding Raf is equivalent to sequestering away some of the activated RasGTP in an “inactive” complex. When Raf concentration crosses a threshold, the bifurcations annihilate each other and disappear. This system is now characterized by a single stable point for all concentrations of SOS, and the disappearance of the threshold for Ras activation. While there appears very little free Ras, in reality, even for low SOS concentrations there is a large concentration of the activated RasGTP-Raf complex (since RasGTP in these complexes is also protected from the action of the Ras GTPases).

This can be seen in another way in Appendix II: Fig. S4.8 where the stable state of RasGTP is plotted against the level of total Raf in the system, keeping the level of SOS constant. Again we see that a bistable system is transformed into a monostable system when Raf increases beyond a threshold. These results are exactly the same for the model which assumes Michaelis-Menten kinetics, except for small changes in molecule numbers, as can be seen in Appendix II: Fig. S4.7 and S4.9. Results do not change on changing load-binding parameters (Fig. S4.10, S4.11)

Thus the addition of the Raf scaffold, which is an integral part of the MAPK cascade, fundamentally changes the qualitative behavior of the positive feedback switch. The main reason why the steady state bifurcation properties are affected here in contrast to the basic genetic toggle switch is that for this signaling circuit, as seen in Eq. 22–25, total Raf and Ras are conserved, as is typical for a short timescale signal transduction system. These conservation laws couple Raf concentration to RasGTP concentration even at steady state. Therefore adding Raf to the system effectively reduces total Ras concentration since Raf sequesters away Ras from the switch.

To see this more generally, consider for example a chemical reaction system comprising of n-species $Y_1,\ldots Y_n$. Let us assume without loss of generality that the species Y_n is coupled to a downstream circuit through a binding reaction with a load, L. The (n+2) differential equations describing this system are:

$$\frac{dY_1}{dt} = f_1(Y_1,\ldots Y_n) \tag{26}$$

$$\frac{dY_n}{dt} = f_n(Y_1,\ldots,Y_n) + k_{on}[Y_nL] - k_{off}[Y_n][L] \tag{27}$$

$$\frac{d[L]}{dt} = k_{off}[Y_nL] - k_{on}[Y_n][L] \tag{28}$$

$$\frac{d[Y_nL]}{dt} = k_{off}[Y_nL] - k_{on}[Y_n][L] \tag{29}$$

Note that for simplicity of notation we have not indicated the dependence of the dynamical system on its own parameter values. Now in the steady state, if the set of equations is complete, the left side uniformly goes to zero and we recover the result that the steady state remains exactly the same with or without a load, as for the genetic toggle switch. However let us now assume that we have an additional conservation law, say,

$$Y_n^{(0)} = Y_n + [Y_nL] \tag{30}$$

This conservation law implies that one equation in our dynamical system is redundant, and we need to drop one equation to make the system linearly independent. We can decide to drop Eq. 19, and substitute $Y_n = Y_n^0 - [Y_nL]$ in Eq. 20 and Eq. 21 and solve the resulting (n+1) equations for the (n+1) unknowns, $Y_1,\ldots,Y_{n-1},Y_nL,L$, obtaining Y_n as a residual from Eq. 22. Thus the steady state solutions of the $Y_i's$ now involve the amount of the load. Clearly, the existence of the conservation law has led to a change in the steady state properties of the dynamical system. Note that Y_n itself would usually enter (by itself or in the form of other complexes, which then would also need to be accounted for in the conservation law Eq. 22) into one or more of the equations for the remaining species, $Y_1,\ldots Y_{n-1}$. This would result in the equations for those other species explicitly involving, and thus depending upon the level of the load. For the Ras system above, Eq. 16 couples the load, Raf, to the concentration of Ras. However Ras concentration and SOS concentration are also coupled. Thus the load explicitly affects the steady state values of all species concentrations in this system. This leads to a fundamental qualitative change in the bifurcation properties of the system.

4.5 Discussion

It has been pointed out previously that significant sequestration effects can abrogate zero order ultrasensitivity [26], [47], [48], can change the dynamics of simple phosphorylation circuits [23], [24] and change oscillatory behavior in some circuits [27]. We add to this body of work by demonstrating that the addition of a simple binding partner to the output protein of a genetic or signaling switch can have dramatic effects on its properties, and can fundamentally change the operation of the switch.

For a genetic toggle switch with two mutually repressing proteins such as the classic switch built by Gardner et al. [5] we showed that even though the presence of the binding partner does not alter steady state properties of the switch, it can drastically change the dynamic properties. Using a novel potential landscape analysis, we showed that this is because the addition of the binding partner skews the underlying quasi-potential, making one state significantly more stable than the other. In practice therefore, a genetic toggle switch that is significantly skewed towards one side may never properly function as a switch. Thus the downstream consequences of such loads need to be taken into account when designing larger synthetic circuits with the toggle switch as one of the elements.

On the other hand this phenomenon actually provides a way of making artificial switches tunable. It is possible to engineer a biased switch merely by adding a load on the opposite side of the toggle, which is a useful device when engineering a switch that is designed to be switched on only in special circumstances. A load on both repressor proteins similarly stabilizes both sides of the toggle switch. This could be useful when working with synthetic components with low concentrations in cells, especially those that display stochastic switching. A load on both repressor proteins can significantly increase the stability of such a toggle.

In natural systems, mutually repressing toggle switches are often found with other complexities, such as a positive feedback motif on one side. The positive feedback motif by itself biases the toggle switch by stabilizing the side it is on at the expense of the other side. A load on the same side then stabilizes the opposite side, and can re-establish balance between the two quasi-potential wells. For engineering circuits in multi-cellular organisms, it is worth noting that that feedback between the load on a toggle switch and the strength of the positive feedback may ensure that the switch operates efficiently even in the presence of cell to cell variability in the load. How loads vary between cells and in multi-cellular organisms is an interesting question to explore in future work. The presence of the positive feedback provides a potential target for evolutionary fine-tuning of the switch.

In the above analyses we use novel potential landscape methods that have proved useful and insightful in fields such as protein folding to discuss the fundamental properties of a dynamical system that shows not apparent changes in its stability properties. We demonstrate that these methods, though still relatively underdeveloped for use with non-equilibrium chemical reaction systems, hold promise for understanding the dynamics of such systems beyond what linear stability analysis can provide. However there are certain conditions when addition of a load changes the stability properties of the genetic toggle switch. One class of such effects happen when the repressor can leak away from the repressor-load complex, as can happen either when the repressor can decay or degrade when bound to the load, or when the load can modify the repressor and make it unable to repress. We show, employing standard bifurcation analysis, that additional loads in this system can abrogate the switch-like properties of the toggle switch entirely.

In switches based on autocatalysis or positive feedback with an enzymatic deactivation, such as is often found in signaling systems, the effects of a load are equally dramatic. We show that in a simple model of Ras activation, adding a small concentration of Raf molecules changes the bifurcation diagram of the signaling circuit and can completely abrogate the bistability in the system. While we have chosen a specific example of Ras activation, our simplified model, with an autocatalytic forward reaction and an enzymatic backward reaction is a minimal model for a many positive feedback switches. The change in the bifurcation diagram arises from the conservation laws that couple the concentration of the load with the concentrations of the proteins in the upstream module. Given the sensitivity of non-linear dynamical systems to initial conditions, it should probably be expected that many, if not all, positive feedback based switches that operate at the short timescales of signal transduction, and therefore must possess these conservation laws, should exhibit this sensitivity to the effect of a load.

Our results throw up an interesting puzzle for quantitative biologists. In many natural signal transduction systems such as the MAPK cascade, the concentration of the output of a bistable switch is quite comparable to the concentration of the load, thus significant changes in load concentrations could have dramatic effects on the behavior of the switch. However it has also been shown that there is a significant cell to cell variability in protein concentrations [49]. How do cells ensure that positive feedback based switches such as the Ras switch continue to operate robustly in the bistable regime? Additional regulatory mechanisms involving feedback between the load and its partner protein may exist that confer robustness to the qualitative behavior of the biochemical switch. Arguably some of the bells and whistles of natural protein networks that are often disregarded when analyzing the network may in fact be performing this role. In other words, self-assembled switches have to be complex! In this context it is worth mentioning that it has been persuasively argued [50], [51] that some biological circuits maintain robustness of “fold-change’ behavior rather than absolute levels of protein concentration. It is possible that additional protein-protein interactions that couple concentrations of loads with output proteins may end up in performing this function. Another significant factor that needs consideration is the role of spatial segregation in producing feedback from the downstream module to the upstream one. In fact it has been shown experimentally that MAPK substrates sequester activated MAPK in the nucleus, and thus protect it from cytoplasmic phosphatases. Changing the concentration of one substrate therefore affects the concentration of activated MAPK [52].

Previous discussions of the effect of loads on the operation of circuits have suggested the use of insulators, that is circuit elements that insulate the upstream module from the downstream module [22]. The initial suggestions for building insulators in Ref. [22] involved incorporating signal amplification along with negative feedback in the upstream circuit. Another way of insulating the circuit is to ensure that the demand of the load for its cognate repressor is never significant compared to the total amount of repressor. For a genetic switch therefore, a possible insulating mechanism is if the link to the downstream circuit is through a promoter. For example, consider making an AND gate from an output of the toggle switch. This can be done by inserting a constitutively produced protein Y that binds to R1 such that the complex is a transcription factor for another protein, say Z. Thus there is an AND relationship between the two inputs, Y and R1 and the output Z. To offset the effect of load induced modulation of the dynamics of R1, an additional step can be inserted such that R1 first binds to the promoter region of another gene that codes for protein X and activates its transcription, and it is the protein X, rather than R1, that can bind to Y and activate production of Z. The advantage of adding this extra step is that the concentration of the promoter for X is very small compared to the concentration of R1, and therefore load induced modulation of the upstream toggle can be kept at a minimum. Note however that this cannot be done without the additional cost of the time delay required for the transcription and translation of X.

As can be seen, any additional step or series of steps that can amplify a weak signal can act as an insulator. Another standard example of an amplifying circuit is a phosphorylation cascade which is especially relevant when considering Ras activation since it directly leads to the MAPK phosphorylation cascade. Phosphorylation cascades are also very fast, and therefore do not face the additional time delays of an additional transcriptional step. From the point of view of synthetic circuit design, the insulating mechanism here could be constructed by designing a weak binding affinity of Ras (or the synthetic protein that plays that role) for Raf (or the equivalent protein). The bound complex then catalyzes a phosphorylation cascade that ends by connecting to the downstream circuit.

Note that this method of insulation does not have the same time delay costs as the additional transcription steps. However it does come with the metabolic costs of having to produce large amounts of proteins that are essentially serving no useful physiological purpose for the cell. This cost could be relevant in some synthetic biology applications, and certainly needs to be evaluated during circuit design. It has been shown in the context of phosphorylation cycles that insulation always carries a metabolic cost, and in general better insulation carries a greater metabolic cost [53].

The existence of the MAPK phosphorylation cascade however begs the question whether it serves the purpose of insulation of the upstream Ras circuit from the downstream circuit. While it is not possible to answer this intriguing question without further experiments, it does appear that the Ras-Raf complex is present is quite large numbers on activated cells. This would suggest that insulation is not the function for which the cascade may have evolved. Our own analysis of the genetic toggle switch with the positive feedback motif suggests that Nature may prefer more complicated forms of regulation that balance the different components of the circuit. However there is no reason why both methods cannot be utilized. To our mind this is a very exciting question that requires more attention from experimentalists and theorists alike.

It should also be noted that due to non-specific binding of transcription factors with DNA as well as between proteins, every circuit in the cell, real or synthetic, operates in the presence of a load. Variability in the functioning of circuits that are seen when transferring synthetic circuits between species, or even in different cells, may be a result of not only differences in basic protein concentrations, but also of this undervalued but nevertheless tangible load. Based on this reasoning we predict that some of the host-dependent effects that complicate synthetic biology, i.e. a synthetic circuit that works in one organism not performing well in another, are in fact due to changes in the intrinsic load due to non-specific binding when changing hosts.

Our analysis underscores the importance of incorporating loads when simulating models of switches in natural and synthetic systems. Mathematical analysis of switch-like motifs therefore would do well to at least include a load on their output proteins, in order to incorporate the possible effects of load induced modulation on the circuit.

REFERENCES

[1] Wagner GP, Pavlicev M, Cheverud JM (2007) The road to modularity. Nature reviews Genetics 8: 921–931.

[2] Cooling MT, Rouilly V, Misirli G, Lawson J, Yu T, et al. (2010) Standard virtual biological parts: a repository of modular modeling components for synthetic biology. Bioinformatics 26: 925–931.

[3] Prasad A (2012) Computational Modeling of Signal Transduction Networks: A Pedagogical Exposition. In: Liu X, Betterton M, editors. Computational Modeling of Signaling Networks: Springer.

[4] Basu S, Mehreja R, Thiberge S, Chen M-T, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. Proc Natl Acad Sci U S A 101: 6355–6360.

[5] Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403: 339–342.

[6] Chang DE, Leung S, Atkinson MR, Reifler A, Forger D, et al. (2010) Building biological memory by linking positive feedback loops. Proceedings of the National Academy of Sciences of the United States of America 107: 175–180.

[7] Kramer BP, Viretta AU, Daoud-El-Baba M, Aubel D, Weber W, et al. (2004) An engineered epigenetic transgene switch in mammalian cells. Nat Biotechnol 22: 867–870.

[8] Ham TS, Lee SK, Keasling JD, Arkin AP (2008) Design and construction of a double inversion recombination switch for heritable sequential genetic memory. PLoS One 3.

[9] Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. Mol Syst Biol 3: 133–133.

[10] Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’. Nature 469: 212–215.

[11] Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403: 335–338.

[12] Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. Cell 113: 597–607.

[13] Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, et al. (2008) A fast, robust and tunable synthetic gene oscillator. Nature 456: 516–519.

[14] Morey KJ, Antunes MS, Albrecht KD, Bowen TA, Troupe JF, et al. (2011) Developing a synthetic signal transduction system in plants. Methods Enzymol 497: 581–602.

[15] de Lorenzo V (2008) Systems biology approaches to bioremediation. Current opinion in biotechnology 19: 579–589.

[16] Alper H, Stephanopoulos G (2009) Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nat Rev Microbiol 7: 715–723.

[17] Lu TK, Collins JJ (2007) Dispersing biofilms with engineered enzymatic bacteriophage. Proc Natl Acad Sci U S A 104: 11197–11202.

[18] Anderson JC, Clarke EJ, Arkin AP, Voigt CA (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. J Mol Biol 355: 619–627.

[19] Lu TK, Collins JJ (2009) Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. Proc Natl Acad Sci U S A 106: 4629–4634.

[20] Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440: 940–943.

[21] Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nature reviews Molecular cell biology 10: 410–422.

[22] Del Vecchio D, Ninfa AJ, Sontag ED (2008) Modular cell biology: retroactivity and insulation. Mol Syst Biol 4: 161.

[23] Ventura AC, Jiang P, Van Wassenhove L, Del Vecchio D, Merajver SD, et al. (2010) Signaling properties of a covalent modification cycle are altered by a downstream target. Proc Natl Acad Sci U S A 107: 10032–10037.

[24] Jiang P, Ventura AC, Sontag ED, Merajver SD, Ninfa AJ, et al. (2011) Load-induced modulation of signal transduction networks. Sci Signal 4: ra67.

[25] Kim KH, Sauro HM (2010) Fan-out in gene regulatory networks. J Biol Eng 4: 16–16.

[26] Bluthgen N, Bruggeman FJ, Legewie S, Herzel H, Westerhoff HV, et al. (2006) Effects of sequestration on signal transduction cascades. The FEBS journal 273: 895–906.

[27] Jayanthi S, Del Vecchio D (2012) Tuning genetic clocks employing DNA binding sites. PLoS One 7: e41019.

[28] Pomerening JR, Sontag ED, Ferrell JE Jr (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. Nat Cell Biol 5: 346–351.

[29] Prasad A, Zikherman J, Das J, Roose JP, Weiss A, et al. (2009) Origin of the sharp boundary that discriminates positive and negative selection of thymocytes. Proc Natl Acad Sci U S A 106: 528–533.

[30] Das J, Ho M, Zikherman J, Govern C, Yang M, et al. (2009) Digital signaling and hysteresis characterize ras activation in lymphoid cells. Cell 136: 337–351.

[31] Bagowski CP, Ferrell JE Jr (2001) Bistability in the JNK cascade. Curr Biol 11: 1176–1182.

[32] Arkin A, Ross J, McAdams HH (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. Genetics 149: 1633–1648.

[33] Ferrell JE Jr (1996) Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. Trends Biochem Sci 21: 460–466.

[34] Maamar H, Dubnau D (2005) Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop. Molecular microbiology 56: 615–624.

[35] Legewie S, Bluthgen N, Herzel H (2006) Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. PLoS computational biology 2: e120.

[36] Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, et al. (2004) Programmable cells: interfacing natural and engineered gene networks. Proc Natl Acad Sci U S A 101: 8414–8419.

[37] Jayanthi S, Nilgiriwala KS, Del Vecchio D (2013) Retroactivity controls the temporal dynamics of gene transcription. ACS synthetic biology 2: 431–441.

[38] Kuhlman T, Zhang Z, Saier MH Jr, Hwa T (2007) Combinatorial transcriptional control of the lactose operon of Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 104: 6043–6048.

[39] Kim K-Y, Wang J (2007) Potential energy landscape and robustness of a gene regulatory network: toggle switch. PLoS Comput Biol 3: e60.

[40] Wang J, Zhang J, Yuan Z, Zhou T (2007) Noise-induced switches in network systems of the genetic toggle switch. BMC Syst Biol 1: 50–50.

[41] Tian T, Burrage K (2006) Stochastic models for regulatory networks of the genetic toggle switch. Proc Natl Acad Sci U S A 103: 8372–8377.

[42] Warren PB, ten Wolde PR (2005) Chemical models of genetic toggle switches. J Phys Chem B 109: 6812–6823.

[43] Wang J, Xu L, Wang E, Huang S (2010) The Potential Landscape of Genetic Circuits Imposes the Arrow of Time in Stem Cell Differentiation. Biophysical Journal 99: 29–39.

[44] Strasser M, Theis FJ, Marr C (2012) Stability and multiattractor dynamics of a toggle switch based on a two-stage model of stochastic gene expression. Biophys J 102: 19–29.

[45] Allen RJ, Warren PB, Ten Wolde PR (2005) Sampling rare switching events in biochemical networks. Phys Rev Lett 94: 018104–018104.

[46] Wang J, Xu L, Wang E (2008) Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. Proceedings of the National Academy of Sciences of the United States of America 105: 12271–12276.

[47] Buchler NE, Louis M (2008) Molecular titration and ultrasensitivity in regulatory networks. Journal of molecular biology 384: 1106–1119.

[48] Lee TH, Maheshri N (2012) A regulatory role for repeated decoy transcription factor binding sites in target gene expression. Molecular Systems Biology 8: 576.

[49] Raj A, Van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135: 216–226.

[50] Shoval O, Goentoro L, Hart Y, Mayo A, Sontag E, et al. (2010) Fold-change detection and scalar symmetry of sensory input fields. Proc Natl Acad Sci USA 107: 15995–16000.

[51] Goentoro L, Shoval O, Kirschner MW, Alon U (2009) The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol Cell 36: 894–899.

[52] Kim Y, Paroush Z, Nairz K, Hafen E, Jimenez G, et al. (2011) Substrate-dependent control of MAPK phosphorylation in vivo. Mol Syst Biol 7: 467.

[53] Barton JP, Sontag ED (2013) The energy costs of insulators in biochemical networks. Biophys J 104: 1380–1390.

CHAPTER 5: CHANGES IN CELL SHAPE ARE CORRELATED WITH METASTATIC POTENTIAL IN MURINE AND HUMAN OSTEOSARCOMAS [3]

5.1 Summary

Metastatic cancer cells for many cancers are known to have altered cytoskeletal properties, in particular to be more deformable and contractile. Consequently shape characteristics of more metastatic cancer cells may be expected to have diverged from those of their parental cells. To examine this hypothesis we study shape characteristics of paired osteosarcoma cell lines, each consisting of a less metastatic parental line and a more metastatic line, derived from the former by *in vivo* selection. Two-dimensional images of four pairs of lines were processed. Statistical analysis of morphometric characteristics shows that shape characteristics of the metastatic cell line are partly overlapping and partly diverged from the parental line. Significantly the shape changes fall into two categories, with three paired cell lines displaying a more mesenchymal-like morphology, while the fourth displaying a change towards a more rounded morphology. A neural network algorithm could distinguish between samples of the less metastatic cells from the more metastatic cells with near perfect accuracy. Thus subtle changes in shape carry information about the genetic changes that lead to invasiveness and metastasis of osteosarcoma cancer cells.

[3] This work has been published in Biology Open and is reproduced here under the Creative Commons License. Lyons, Samanthe M., et al. "Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas." Biology open 5.3 (2016): 289-299. I am co-first author on the publication with Elaheh Alizade. The work is being presented here in its entirety to maintain the intellectual coherence of the project.

5.2 Introduction

Despite significant advances in treatment of cancer, it remains the leading cause of death in both men and women under 80 years of age in the US [1], with metastasis as the cause of 90% of human deaths from cancer [1, 2]. Understanding and prevention of cancer invasion and metastasis is key in reducing cancer mortality [2]. Multiple studies have pointed out that the acquisition of invasiveness appears to require changes in mechanical properties of cancer cells, which may be linked to the functional properties that are necessary for metastasis [3, 4]. To form successful metastases, tumor cells must navigate a complex, multi-stage process including: detachment from primary tumor, migration to vascular supply, intravasation, survival and transit in blood or lymphatic vessels, extravasation, and successful growth and adhesion in a new site [5]. Metastatic cells have been found to be softer or more deformable than non-metastatic cells in analysis with atomic force microscopy [6-9] and optical lasers [10-12]. In addition to cell deformability, multiple studies have shown that molecules responsible for cell-ECM and cell-cell adhesion interactions, including cadherins and integrins, are down-regulated or altered in cancer cells [13-17]. Cancer cell deformability is linked with invasiveness and can be an indicator of metastatic potential [3, 18-20]. However softness is just one aspect of cellular biomechanics. Cells are active objects and can exert contractile forces on the extracellular matrix; there are some reports that more invasive cells are more contractile [21]. Understanding and identifying altered biomechanical properties of aggressive cancer cells can provide crucial information for assessing the invasiveness of cancer cells.

Direct assays of mechanical changes require fairly complex and expensive instrumentation. However, one can hypothesize that changes in biomechanical properties, including changes in cytoskeletal properties and expression of adhesion proteins, would translate into changes in cell shape. It has been shown previously that changes in gene expression of genes with cytoskeletal function leads to shape changes that can be detected using morphometric characteristics [22]. Cytoskeletal gene expression changes that are signatures of metastatic capacity therefore may be detectable by morphometric analysis. Ability to detect such changes would be of great use in assessing cancer clinically.

One of the gold standards of predicting clinical outcome of cancer is tumor grading which includes assessment of cellular morphology from tumor tissue samples. Tumor grading schemes focus on overt changes in cellular morphology such as mitotic index, degree of nuclear pleomorphism and degree of tumor necrosis [23, 24]. What is not known is whether morphometric parameters of the two-dimensional shape of the cell are sensitive to the changes in cellular properties that accompany the acquisition of invasiveness.

Our paper is based on the hypothesis that subtle changes in cellular properties should manifest themselves in small but detectable morphological changes because of the importance of cytoskeletal changes for acquisition of invasive capacity. The biomechanical changes that accompany the emergence of aggressive tumor cells should be detectable by assaying the changes in shape of these tumor cells. Moreover, the observation of specific changes in shape may be linked with specific genetic changes in cancers.

Anecdotal evidence for the change in cell shape has been well documented. For example, the epithelial to mesenchymal transition (EMT), associated with development of the invasive phenotype in carcinomas [25], is often accompanied by acquisition of a mesenchymal-like elongated spindle morphology [26, 27]. Studies have found that tumors which have formed metastases at the time of diagnosis have significantly higher grades, and thus grossly altered morphology, than non-metastatic osteosarcomas [28]. An understanding of the relationship between cell shape and the invasiveness of the cancer would lead to a deeper understanding of the relationship between carcinogenic transformation and shape regulation and may allow for a more accurate assessment of cancer outcome from cancer biopsies. Assays that can reliably estimate the percentage of potentially invasive cells in a heterogeneous sample of primary tumor cells may be of great value for guiding therapeutics.

We utilized osteosarcoma cell lines due to the high metastatic rate of the cancer. Osteosarcoma (OSA) is the most common primary bone tumor of dogs [29] and humans [30]. OSA has a high rate of metastasis and routinely forms metastases to the lung, often before the primary tumor is diagnosed and more than 80% of human OSA patients have metastases at the time of diagnosis [31-34], most with pulmonary metastases [35, 36].

Comparing the morphology of cells that were closely related (except for their degree of invasiveness) was important to minimize variables that would make the sample less homologous. We therefore sought paired osteosarcoma lines, where a more aggressive cell line was developed from a less aggressive ancestor without the use of exogenous transforming agents, as these agents may alter naturally occurring genetic changes leading to metastatic properties of osteosarcoma [37]. Without exogenous agents, we can attribute changes in cell morphology more directly to the difference in metastatic potential, as the *in vivo* development of the highly metastatic line more accurately represents the natural process of formation of metastases.

The morphology-related genetic changes that accompany transformation include both changes in cytoskeletal properties as well as changes in adhesive properties [38]. We decided to use surfaces of different hydrophobicity in our experiments to explore this possibility as more hydrophobic surfaces are less amenable to protein deposition [39] and thereby are less favorable to cell adhesion than hydrophilic surfaces. We prepared three different glass surfaces of varying hydrophobicity (Appendix II: Supplementary Fig. 5.1). These are Glass Detergent washed and Air dried (GDA, contact angle 27.6°), Glass Acid etched and Air dried (GAA, contact angle too small to measure), and Silconized Ethanol Treated (SET, contact angle 99°).

We cultured four paired osteosarcoma cell lines with low and high metastatic potential: DUNN and DLM8; K12 and K7M2; MG63 and MG63.2; and SaOS2 and SAOS-LM7 on these three surfaces for 48 hours, and then fixed, stained and imaged the cells. For simplicity we refer to each pair by the first letter of the parental line, i.e. we refer to the pairs as the D, K, M and S pairs of lines. We stained the cells for actin, the plasma membrane and nucleus. We developed a high-throughput, quantitative image analysis algorithm that chose individual cells not in contact with others, segmented, optimized and thresholded the images to obtain accurate representations of two-dimensional shape and then processed the images to extract 29 morphometric measurements: 21 cellular and 8 nuclear (Appendix III: Supplementary Table 5.1). Representative images of the eight different cell lines are shown in Fig. 5.1. Since here we are specifically looking for interpretable geometric differences, we did not consider other morphological representations such as shape representations in basis function expansions [40]. We then subjected the data to statistical analysis to understand the differences between the high metastatic and low metastatic cell lines, using pairwise comparisons as well as by the multivariate Principal Component Analysis (PCA) and Nonmetric Multidimensional Scaling (NMDS). We developed a neural network machine-learning algorithm to try to distinguish between cells from the high metastatic and low metastatic cell lines.

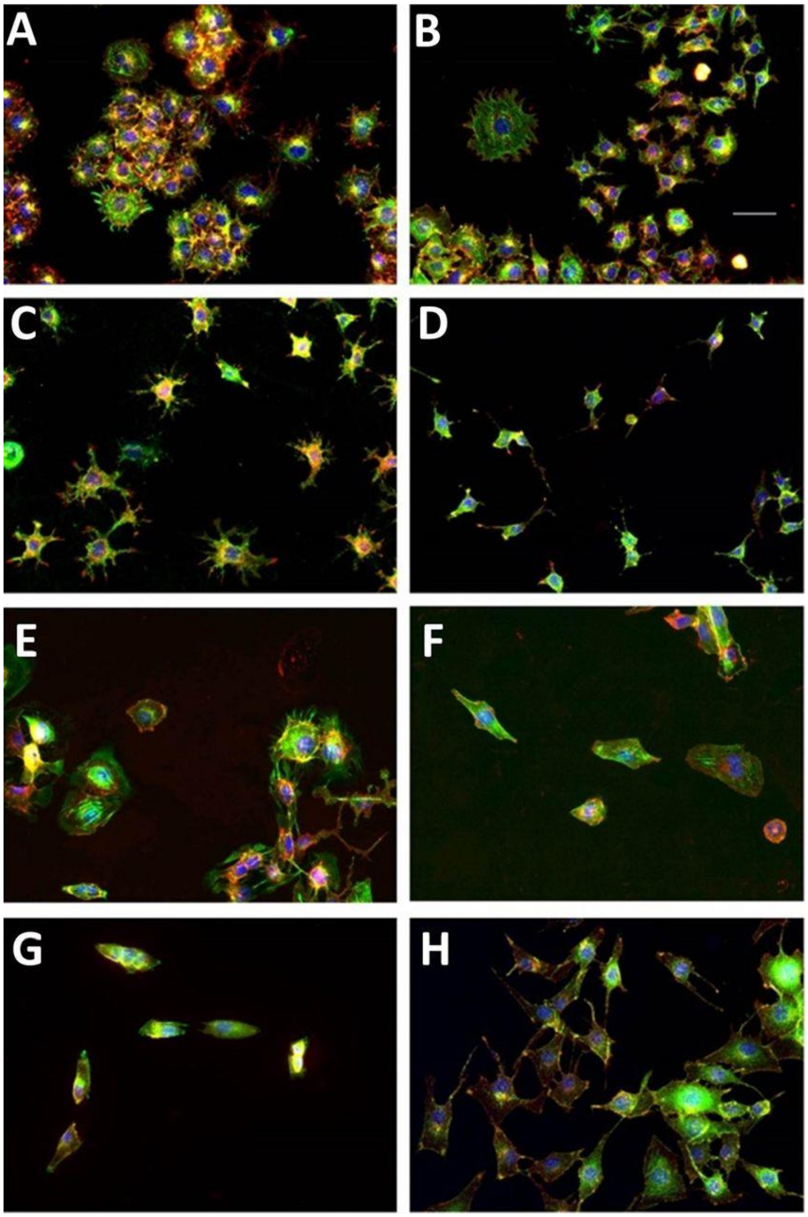


Figure 5.1. Representative images of the four cell lines using fluorescence microscopy. Each set of two panels represent the low metastatic (left) and the high metastatic (right) partner of a paired cell line. The cells nuclei (blue), the actin cytoskeleton (green) and the lipid membrane (red) of fixed cells are stained and are pseudo-colored as indicated for contrast. Note that the yellow color indicates the overlap of the red (membrane) and green (actin) channels. The cell lines are: (A) Dunn, (B) DLM8, (C) K12, (D) K7M2, (E) Saos2, (F) SAOS-LM7, (G) MG63 and (H) MG63.2. In all panels, scale bar is 50 µm.

5.3 Results

5.3.1 Pairwise Comparisons: The four paired cell lines demonstrated two distinct trends of cell shape changes

The 29 morphometric parameters were classified into five categories of cell shape: (i) projected cell size, (ii) cell roundness vs elongation, (iii) shape variability, (iv) nuclear size, and (v) nuclear shape. We identified a subset of the 29 parameters that were most often statistically significant across the various cell lines by performing pairwise t-tests between different morphometric measurements of the low metastatic line (low-met) and high metastatic line (high-met) within a paired cell line. In order to adjust for multiple testing, we performed t-tests on all 29 parameters using the Holm-Bonferroni correction [41], and identified the parameters that remain significantly different (Appendix III: Supplementary Tables 5.2 and 5.3). The data showed that metastatic cell lines show distinct differences in shape compared to their non-metastatic counterparts. Significantly, we discovered that three of the four paired cell lines, i.e. the D, K, and S lines displayed a similar pattern of shape changes, while the fourth line, i.e. the M-line, displayed a different pattern (Fig. 5.2 and Appendix III: Supplementary Table 5.2). This suggests that morphological changes due to acquisition of metastatic potential fall into two distinct classes. For simplicity we denote these two patterns as type-1 and type-2. When pooled into two classes, the type-1 cells showed significant differences in 28 out of the 29 parameters we tested, while the type-2 cells showed significant differences in 26 parameters (Appendix III: Supplementary Table 5.3).

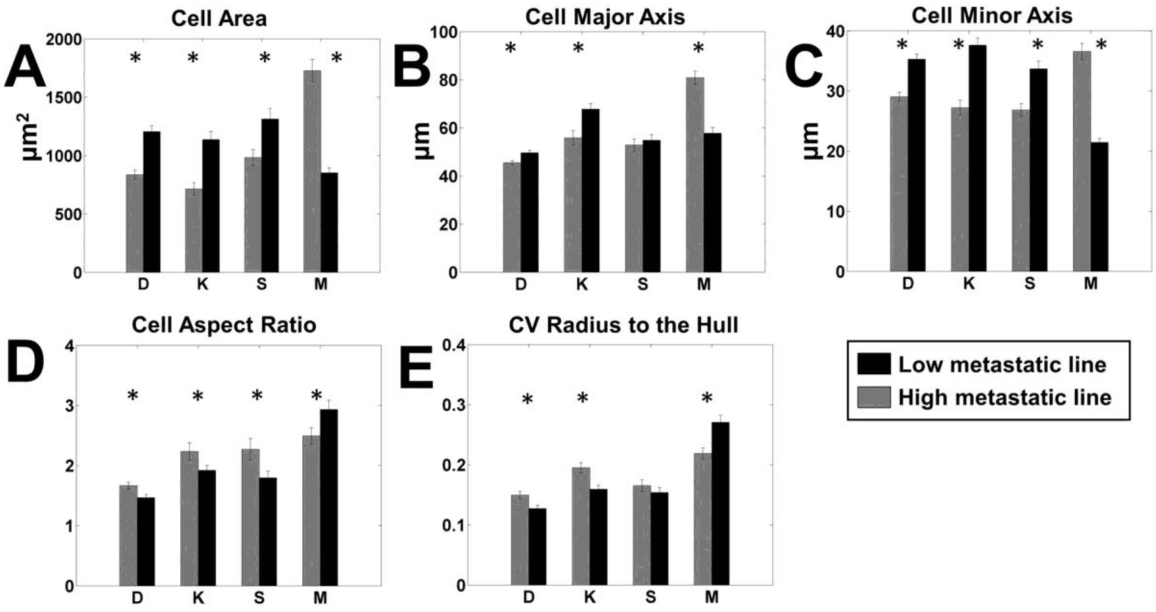


Figure 5.2. Pairwise comparison of most significant cell shape parameters.
Each panel shows the comparison between high metastatic (grey) and low metastatic (black) cell lines for a single significant parameter on all surfaces. The paired lines are indicated by letters as follows. D: DUNN and DLM8; K: K12 and K7M2; S: Saos2 and SAOS-LM7; M: MG63 and MG63.2. (A) Cell area, (B) cell major axis, (C) cell minor axis, (D) cell aspect ratio and (E) coefficient of variation (CV) of the radius from the center of mass to the hull. n=100 for each cell line on each surface. *P<0.05 by two-tailed t-test satisfying the Holm–Bonferroni criteria for all variables (Appendix III: Table S3).

5.3.2 Highly metastatic cancer cells differ in cell volume and projected cell area

A striking gross morphological difference between the high-met and low-met line of each pair is a systematic difference in two-dimensional projected area. Less metastatic type-1 cell lines have a significantly larger projected cell area (Fig. 5.2A), and on average, the type-1 high metastatic lines are 30.7% smaller in area. The type-2 M lines showed the opposite trend, with highly metastatic cells being significantly more spread out, more than double the size of the parental line on some surfaces (Appendix III: Supplementary Table 5.2). To determine whether cell volume corresponded with cell area, we measured cell volume using a handheld Scepter® counter (Methods and Materials). We found that for most pairs the cell volume and the cell area followed the same trend, i.e. the high-met line was smaller in both area and volume for two lines of type I (K and S), while the high-met M line was larger in volume and area than its less metastatic pair. The D line showed an opposite trend with a large volume but smaller area for the high-met line. However the percentage difference in mean volume is much smaller than the percentage change in mean area, suggesting a difference in the spreading behavior of the cells for both type-1 and type-2 cells. This trend of a smaller projected high-met line for type-1 cells and a larger projected high-met line for type-2 cells is consistent across all 12 measures of two-dimensional cell size (Appendix III: Supplementary Table 5.2). Within the type 1 cells, the largest difference was shown by the metastatic K12 cells that were over 50% smaller on GDA while the smallest difference was shown by the LM7 cells which were about 23% smaller (Appendix III: Supplementary Table 5.2).

The change in size is also anisotropic, as the type-1 high-met lines have a minor axis that is 22% smaller, while the major axis is only 10.5% smaller, thus the minor axis percent difference is about twice that of the major axis (Fig. 5.2 B&C), indicating elongation of the high-met type-1 cells relative to the low-met cells. The smaller size and elongated shape of high-met lines in type-1 osteosarcoma pairs are consistent across all three type-1 lines. The type-2 M cell lines showed the opposite trend for the major and minor axes with the high-met line having a larger minor axis by 67% and larger major axis by 48%, respectively.

5.3.3 Highly metastatic cells differ in roundness, elongation and variability of perimeter

The second category of cell shape assesses how round versus elongated the cell is, and is best represented by the aspect ratio. As suggested by the major and minor axis differences discussed above, the type-1 highly metastatic cell lines had a significantly larger aspect ratio than the low-met lines, indicative of cell elongation (Fig. 5.2D). On average, the type-I highly metastatic cells were about 19% more elongated than the low metastatic cells. The maximum change here were the LM7 cells on SET with a 60% increase in the aspect ratio, while the smallest were the DLM8 cells with just about 13% increase on SET surfaces.

Variability of cell shape was characterized by the cell shape parameters of solidity and the coefficient of variation of radii drawn either to the cell perimeter from the center of mass of the convex hull (CV Rad Hull) or from the center of mass of the bounding circle to points on the convex hull (CV Rad Circle). The highly metastatic type-1 cell lines had more variability in radii drawn to both the perimeter (Appendix III: Supplementary Table 5.2 and Fig 5.2E) and the convex hull. Another interesting measure is the circularity of the perimeter, which measures the deviation of the average shape from that of a circle. Circularity of the cell perimeter is significantly different between the high-met and low-met type-1 lines, by a little over 37% on average (Appendix III: Supplementary Table 5.2).

The type-2 M lines showed the opposite trend to the ones listed above. The low metastatic cells had an aspect ratio which was about 22% larger on the GAA surface, and about 15% larger overall. The type-2 low-met line also displayed greater variability in shape than the high-met line with the CV of the perimeter radius larger by about 19% on average and by about 25% on the GAA surface. Similarly the CV of the Hull radius was larger by almost 28% on average for the low-met line. The circularity of the low-met type-2 line was also larger than its high-met partner, in contrast to the behavior shown by the type-1 lines.

5.3.4 Highly Metastatic Cell Lines Show Shape Differences in the Nucleus

Interestingly, the shape parameters of the nucleus also showed statistically significant differences between the high and low metastatic lines (Fig. 5.3). Nuclear size was larger for the low metastatic cells for all cell lines, including both type 1 and type 2. However while the larger nuclear size for low-met cells was statistically significant for the type-1 cells on GAA and SET surfaces, as well as for all surfaces, it was significant for the type-2 line only on the GAA surface. In line with the difference in nuclear area, the major and minor axes were larger for the low metastatic cells for all four pairs of cell lines (Fig. 5.3 B&C). However, the nuclei aspect ratio showed mixed results, with the high metastatic lines demonstrating a larger aspect ratio for the D and M lines, while the low metastatic lines demonstrated a larger aspect ratio for the K and S lines (Fig. 5.3D).

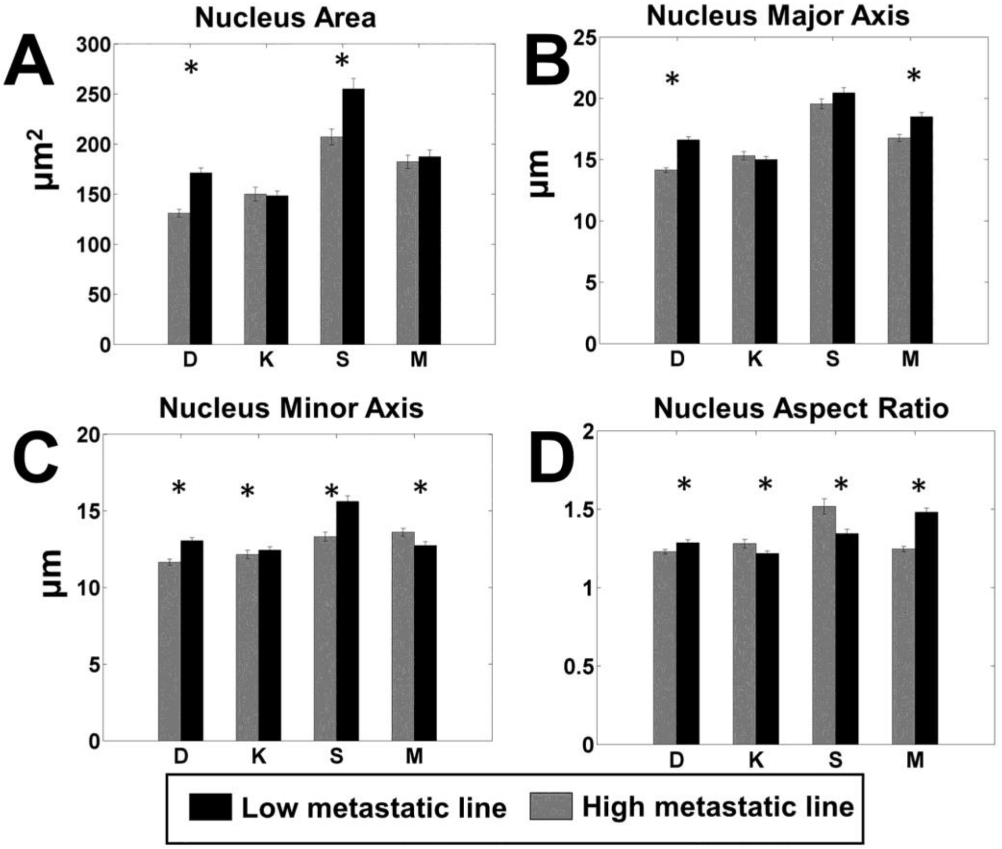


Figure 5.3. Pairwise comparison of the most significant parameters of nucleus shape. Each panel shows the comparison between high metastatic (grey) and low metastatic (black) cell lines for a single significant parameter. The paired lines are indicated by letters as follows. D: DUNN and DLM8; K: K12 and K7M2; S: Saos2 and SAOS-LM7; M: MG63 and MG63.2. (A) Nuclear area, (B) major axis of the nucleus, (C) minor axis of the nucleus and (D) aspect ratio of

the nucleus. n=100 for each cell line on each surface. *P<0.05 by two-tailed t-test satisfying the Holm–Bonferroni criteria for all variables (Appendix III: Table S3).

This analysis also underscored the fact that every cell line contained a heterogeneous collection of cell shapes. The distributions of each parameter overlapped, which was not surprising given the fact that we chose the paired lines on the grounds that they were close to each other in genetic space. In the light of these results, we asked whether we could still see these differences using a multivariate measure by utilizing all the descriptors together.

5.3.5 Multivariate Techniques show overlapping but distinct cell populations

We performed a principal component analysis (PCA) of the multivariate data, comparing each paired line separately (Fig. 5.4 and Appendix III: Supplementary Fig. 5.3). The PCA showed that the geometric characteristics of each cell type were overlapping but clustered distinctly within the space formed by the first three principal components. The overlap between the characteristics of the paired cell lines indicates that the high-met line is still not too dissimilar from the low-met line. However the genetic changes that accompany the acquisition of invasive characteristics have also resulted in the cell shape parameters drifting away from that of the original cell. The maximum overlap of the first three principal components can be seen in the SAOS-LM7 and Saos2 pair (Fig. 5.4D). The type-1 cells collectively show distinct clustering of the low-met and high-met populations (Fig. 5.4E), which is lost when we club the type-1 and type-2 cells together (Fig. 5.4F).

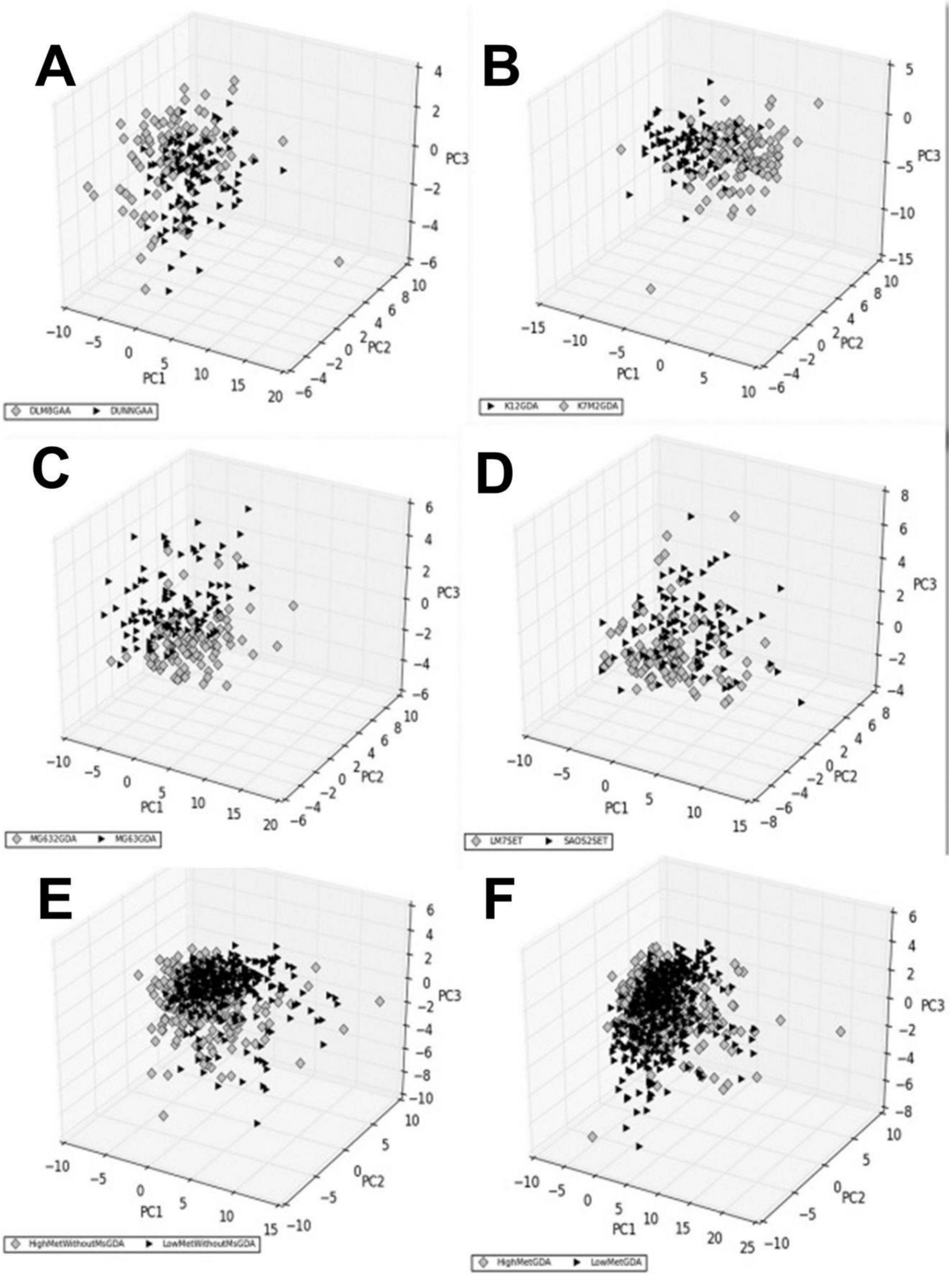


Figure 5.4. Principal components of shape characteristics.
The shape characteristics data for each cell is projected onto the first three principal components of the combined data of each comparison. In this figure, comparisons for each paired cell lines that performed best are shown, as determined by visual inspection and global comparisons. The grey diamonds represent the high metastatic cell line(s) while the black triangles represent the low metastatic line(s). Each panel represents one comparison as follows: (A) DUNN vs DLM8

on GAA, (B) K12 vs K7M2 on GDA; (C) MG63 vs MG63.2 on GDA; (D) Saos2 vs SAOS-LM7 on SET; (E) all type-1 low metastatic lines versus high metastatic lines on GDA and (F) all low metastatic versus high metastatic (i.e. both types combined) on GDA. n=100 for each cell line on each surface.

To test whether we could obtain a better separation using a nonlinear technique, we supplemented PCA by non-metric multidimensional scaling (NMDS) [42]. NMDS is an ordination technique that seeks to find the "best" coordinates for representing multivariate data in a lower k-dimensional ordination space. It does so by assessing and optimizing the agreement between ranked distance between data vectors in the original higher-dimensional space and the corresponding distance between them in k-space. Departure from this agreement is formally measured as "stress". Other groups have used multidimensional scaling (MDS) to visually separate subpopulations of mesenchymal cells, further using this analysis to predict the fate of differentiating stem cells [43]. We used permutational multivariate analysis of variance to obtain the R^2 values, where in the NMDS context R^2 is a measure of the proportion of the distance variation of the data that is explained by cell line, i.e. from the high-met or low-met comparison within each paired line. Fig. 5.5 shows the NMDS results for the best-performing surfaces, and shows that the geometric characteristics overlap between paired lines but nevertheless cluster distinctly. The R^2 values are tabulated in Table 5.1 along with their p-values. The maximum proportion of the distance variation that can be attributed to cell line is 0.16 for the D-lines on the GAA surface, 0.2 for the K-lines and 0.06 for the S-lines (both on the GDA surface), and 0.24 for the M-lines on the SET surface (0.22 on GDA). All the R^2 values are statistically significant and indicate that cell shape parameters of the high-met line, despite significant overlap, have diverged from those of the low-met line. Other surfaces show varying levels of overlap but in general support this conclusion (Appendix III: Supplementary Fig. 5.4). Interestingly data points corresponding to the high-met line for both type-1 and type-2 cells occupy a greater area in 2-space, suggesting that the high-met lines are characterized by greater heterogeneity of the shape parameters.

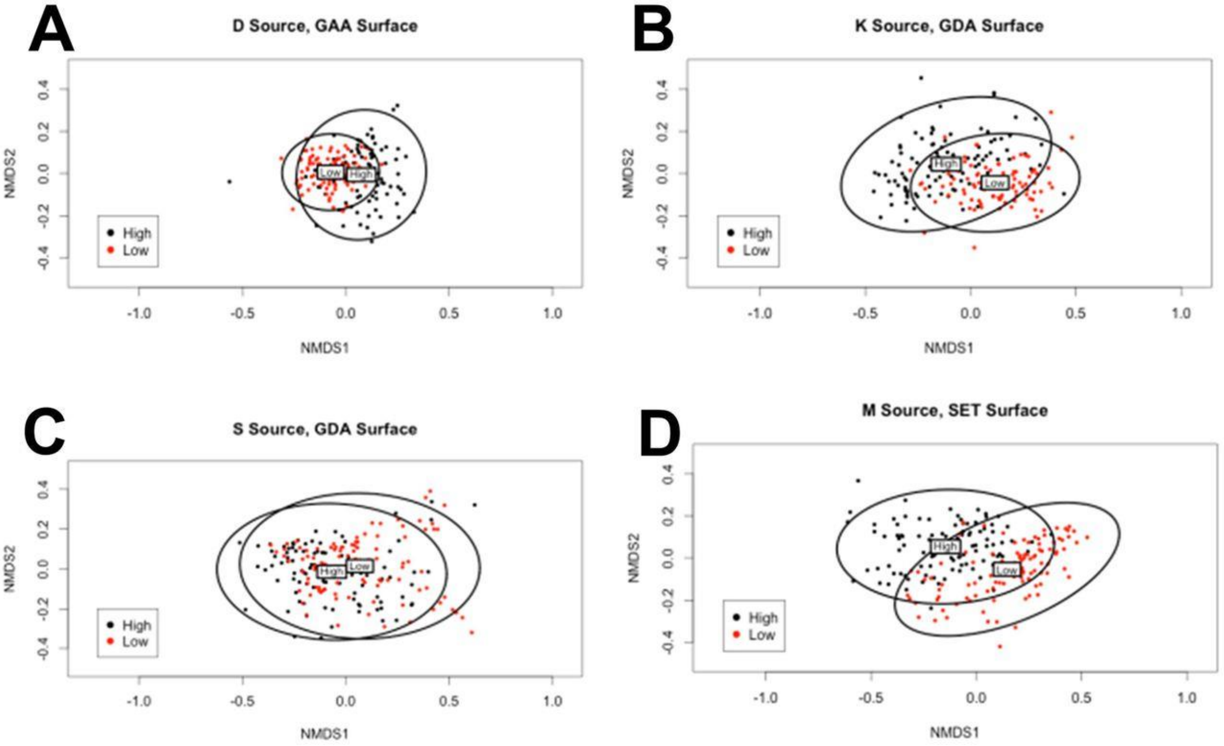
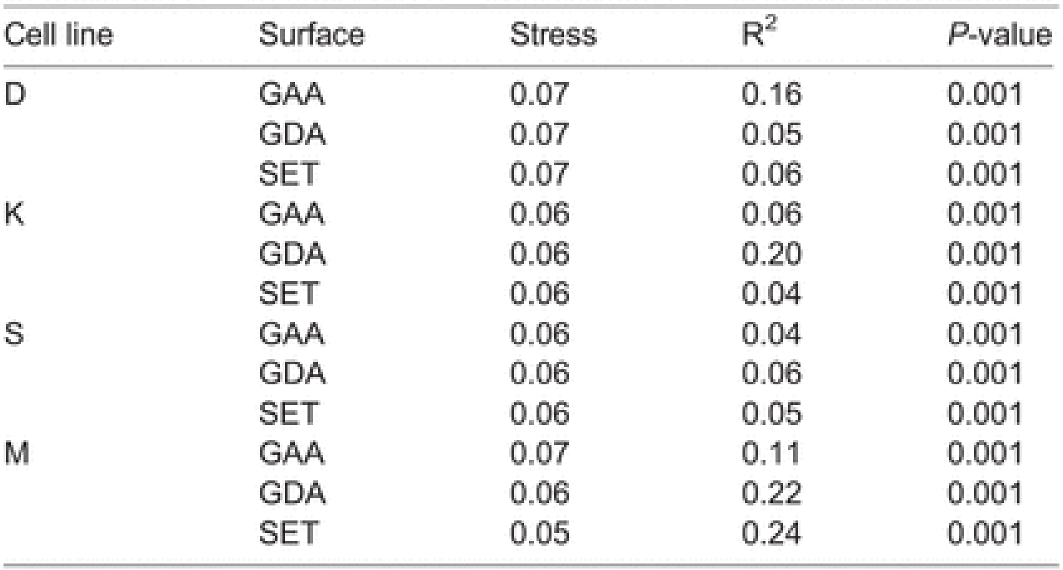


Figure 5.5. Nonmetric multidimensional scaling.
Each panel represents an ordination pattern formed by comparison of geometric characteristics of a low metastatic and a high metastatic cell line on the surface that showed the highest R2 value for the pair. Each point represents the shape parameters of a single cell, plotted in black if high metastatic and red if low metastatic. The ellipses represent 95% confidence intervals with the labels ‘High’ and ‘Low’ marking the centroid positions of the corresponding cell line. The comparisons are as follows: (A) DUNN (low) and DLM8 (high) on GAA; (B) K12 (low) and K7M2 (high) on GDA; (C) Saos2 (low) and SAOS-LM7 (high) on GDA and (D) MG63 (low) and MG63.2 (high) on SET. n=100 for each cell line on each surface.

Table 5.1. Nonmetric multidimensional scaling statistics. Comparisons between a paired cell line on each surface. The 'Stress' is a measure of the departure of the ranked distances of the cells in the low-dimensional NMDS space from that in the original high dimensional space. The low numbers in the table indicate that NMDS was able to preserve the ranked differences. The R^2 (for NMDS) is an average measure of the proportion of the total distance between cells that can be explained by the membership in the two lines, i.e. high-met and low-met. Both the R^2 values and their P-values were calculated using permutation multivariate analysis of variance. The surface abbreviations are as follows: GAA, glass acid etched; GDA, glass detergent washed; SET, siliconized glass, ethanol treated. n=100 for each cell line on each surface.



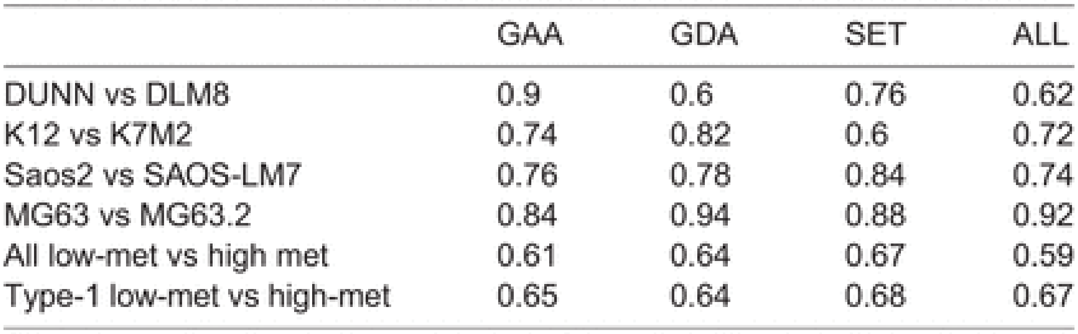
Cell line	Surface	Stress	R^2	*P*-value
D	GAA	0.07	0.16	0.001
	GDA	0.07	0.05	0.001
	SET	0.07	0.06	0.001
K	GAA	0.06	0.06	0.001
	GDA	0.06	0.20	0.001
	SET	0.06	0.04	0.001
S	GAA	0.06	0.04	0.001
	GDA	0.06	0.06	0.001
	SET	0.06	0.05	0.001
M	GAA	0.07	0.11	0.001
	GDA	0.06	0.22	0.001
	SET	0.05	0.24	0.001

5.3.6 Identification of Cells by Machine Learning

We asked whether these subtle but significant differences in cell shape are sufficient to construct a classification algorithm that could correctly classify the low-met and high-met cells. We wrote a neural network machine-learning algorithm to classify a cell into either the low-met or the high-met class, based on its geometric parameters alone, as described in the Methods section. Following standard practice we divided our data into three mutually exclusive subsets for training, optimizing and validating the neural network respectively. The trained algorithm was then tested blind on the third subset, the validation set, which was not used for any parameter adjustment

The accuracy of classification of the algorithm was found to lie between 60% and 92%, (Table 5.2) suggesting as high as about 40% and as low as 8% overlap of parameters. The latter figure is much lower than expected from the preceding analysis, probably due to the efficiency of the neural network in picking up subtle differences. Single cells from all the four lines can be classified with at least 80% accuracy on at least one surface.

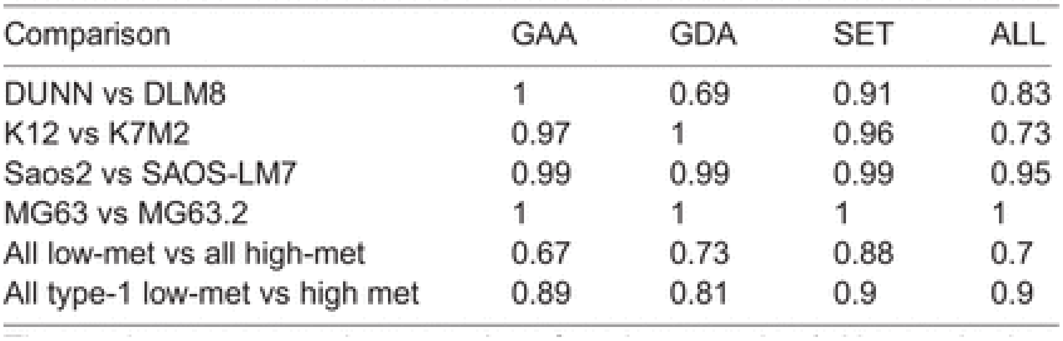
Table 5.2. Proportion of individual cells correctly identified by the neural network algorithm. The numbers represent the proportion of the sum of true positives and true negatives to all cases (see Materials and Methods section). Each row is a specific indicated comparison while the columns represent the surface on which the cells were cultured, with the last column representing the results of data from all surfaces combined. The surface abbreviations are as follows: GAA, glass acid etched; GDA, glass detergent washed; SET, siliconized glass, ethanol treated.



	GAA	GDA	SET	ALL
DUNN vs DLM8	0.9	0.6	0.76	0.62
K12 vs K7M2	0.74	0.82	0.6	0.72
Saos2 vs SAOS-LM7	0.76	0.78	0.84	0.74
MG63 vs MG63.2	0.84	0.94	0.88	0.92
All low-met vs high met	0.61	0.64	0.67	0.59
Type-1 low-met vs high-met	0.65	0.64	0.68	0.67

Next we asked whether the classification algorithm is capable of accurately classifying random samples of cells from the high-met and the low-met lines. This process can be construed as a simulation of what would happen in a clinical setting: the heterogeneous cancer cell population taken from a tumor biopsy or aspirate would be assayed using morphometric characteristics. The decision algorithm used was that if the majority of cells in the sample are of type A, the sample is of type A, and with this simple rule the algorithm achieves near perfect classification of samples into the correct cell type (Table 5.3). For every line there is at least one surface where samples can be classified with greater than 95% accuracy. Even for the S-line, where NMDS revealed only a 6% maximum difference between the cell line parameters, the neural network achieves a maximum classification accuracy of 99%.

Table 5.3. Accuracy in sample identification of the neural network. The numbers represent the proportion of random samples (with sample size 10) from the high metastatic and low metastatic cell lines that were correctly identified by the neural network algorithm. The accuracy is the proportion of the sum of true positives and true negatives to all cases (see Materials and Methods section), hence the maximum possible accuracy is 1. Each row is a different comparison as specified, and the columns represent the three surfaces separately and combined (last column). The surface abbreviations are as follows: GAA, glass acid etched; GDA, glass detergent washed; SET, siliconized glass, ethanol treated.



Comparison	GAA	GDA	SET	ALL
DUNN vs DLM8	1	0.69	0.91	0.83
K12 vs K7M2	0.97	1	0.96	0.73
Saos2 vs SAOS-LM7	0.99	0.99	0.99	0.95
MG63 vs MG63.2	1	1	1	1
All low-met vs all high-met	0.67	0.73	0.88	0.7
All type-1 low-met vs high met	0.89	0.81	0.9	0.9

Note that the algorithm performs relatively poorly when used to classify samples from all low-met lines against all high-met lines as compared with when it is used on type-1 cells and type-2 cells separately. Thus, shape changes in the three paired lines in type-1 appear similar enough that despite originating from different species and different cell lines, they can be accurately classified into high-met and low-met cells with reasonable accuracy.

5.4 Discussion

We have shown that highly metastatic osteosarcoma cell lines derived from less metastatic parental cells show differences in shape that can be broadly classified into two types. In type-1, displayed by 3 out of 4 paired cell lines, the highly metastatic cells are smaller in two-dimensional area, more elongated, and the radius of the cell perimeter from the center of mass is more variable. In type-2 cells, displayed by one cell line pair, the cells become larger, more rounded, with a less variable perimeter. In both cases, the distribution of the geometric characteristics that we measured was more diverse for the high-met cell line. There was a significant overlap between the parameters of the low-met and the high-met line. However use of multivariate data analysis techniques such as PCA and NMDS indicated that despite the overlap, the data points of the two cell lines clustered slightly differently from each other. The differences were sufficient to enable a trained neural network to correctly classify an individual cell as belonging to the high-met or low-met line with over 80% accuracy on at least one surface, and to almost perfectly classify samples of cells from either line.

Our data suggest that genomic changes leading to acquisition of invasive properties also give rise to detectable shape changes, and hence shape changes carry information about the state of the cell. While this study was restricted to these four pairs of osteosarcoma cell lines, we suspect that the broad conclusions are more general. Genetic changes that drive the acquisition of invasive properties may affect cell shape in various ways. For example, there could be down-regulation of adhesive proteins, a softening of the cell due to down-regulation of keratin and up-regulation of vimentin and changes in cellular contractility due to Rho-ROCK signaling. Each of these is likely to have a different set of effects on cell shape, and requires further investigation. Identifying and understanding the full typology of shape changes could have a major impact on our knowledge of metastasis and its relation with the cellular cytoskeleton. It may be eventually possible to read out genetic changes corresponding to specific changes in cell shape [22]. Determining the causal links between genetic changes and the shape of the cell are outside the scope of the present paper (and are future goals), but we provide evidence that these links exist since functional changes in invasive properties correlates with changes in cell shape.

Our discovery that shape changes fall into two types or classes is also potentially significant. Our hypothesis arising from this work is that these two classes correspond to the two modes of cell migration, i.e. mesenchymal and amoeboid [44]. Mesenchymal motion consists of cell polarization, extension, substrate binding followed by actin-based contraction and release of focal adhesions at the trailing edge. This kind of migration is dependent upon adhesion receptors, as well as on the expression of enzymes that degrade the extracellular matrix such as MMPs [44]. However when enzyme activity of MMPs is blocked, cells are found to move in an amoeboid manner, wherein the cell squeezes itself into the empty spaces in the extracellular matrix. The two modes of motion are associated with different morphologies, with the mesenchymal mode corresponding to an elongated morphology and the amoeboid mode corresponding to a rounded morphology [45]. Recent studies have shown that the amoeboid mode of migration is associated with Rho signaling through ROCK and requires the protein ezrin, which links the cell membrane and the cellular cytoskeleton [46]. Thus downregulation of MMPs and upregulation of ezrin appears associated with amoeboid motility. The highly metastatic MG63.2 line was found to be characterized by downregulation of MMPs and upregulation of ezrin [37], suggesting that its preferred mode of motility could be amoeboid and providing an explanation for the rounded morphology it possesses as compared with the parent MG63 line. This suggests that cancer cells may acquire intrinsic preference for one mode of motility over the other as they acquire invasive characteristics, even if they are capable of switching modes of motility [44, 47]. Thus type-1 cells could have an intrinsic preference for mesenchymal motility while type-2 cells could have an intrinsic preference for amoeboid motility.

Our work suggests that it is possible to develop a consistent reproducible framework for computational morphometrics of cell shape. Further work is required to validate and refine the framework through use of other cell lines, including primary tumor lines, other cancer types and species. A reproducible quantitative framework is important for improving the subjectivity of traditional morphological analysis performed by trained histopathologists. While there is a strong correlation between tumor grade and metastatic outcome, there is not yet an ability to predict metastatic potential, based on tumor grade, in individual cases [28]. One study found low reliability of the grading of chondrosarcomas, despite the fact that grading scores guide therapeutic decision-making [48]. A summary of numerous studies on the reliability and reproducibility of urologic, prostrate or renal cell cancer grading found low agreement and reproducibility. [49]. Our work provides some evidence that computational image processing based morphometry to assess tumor grade may help overcome some of these challenges.

A small number of recent publications have highlighted the functional importance of cell shape by using high throughput image analysis to characterize the relation between cellular morphology and cellular properties. Treiser et al. [43] used quantitative morphometric descriptors along with MDS to predict differentiation of mesenchymal stem cells along bone or fat lineages at an early time point. They showed that subtle genetic differences between cells proceeding down the two lineages could be inferred from looking at small changes in cellular morphometrics. Yin et al. [50] utilized high throughput imaging and computational methods to classify *Drosophila* haemocyte cells into 5 discrete shapes based upon quantitative shape and morphology metrics, and argued that transitions between these shapes are switch-like. They utilized RNAi to identify genes which play a large role in regulating cell shape, including demonstrating that the loss of PTEN induces elongation of cells. They did not however look for systematic differences between closely related cancer cell lines. While we have not tried to ascertain whether specific types of shapes are present in our data, the message of this paper is that differences in quantitative shape parameters, even within the same type, should carry useful information about the internal state of the cell. The overlap between the multidimensional shape parameters in principal component space or in NMDS space indicates that each pair of the cell lines we study has not diverged significantly in shape characteristics. However both these studies support our contention in this paper that the understanding of cell shape can give significant insight into cell properties and function.

Studies of metastasis in cancer cells have focused mainly on changes at the level of gene, protein and microRNA expression, and to a smaller extent, at the level of cellular mechanics. In contrast our work demonstrates that these changes do lead to reproducible changes in shape. More work needs to be done to construct a more comprehensive typology of shape changes in cancer, especially in other cancer types, and to achieve a mechanistic understanding of how changes in gene and protein expression result in changes in cell shape.

5.5 Materials and Methods

5.5.1 Cell Lines and Cell Culture

We utilized four paired cell lines; two of murine origin: DUNN and DLM8, K12 and K7M2, and two of human origin: MG63 and MG63.2, SaOS2 and SAOS-LM7. All metastatic lines (DLM8, K7M2, SAOS-LM7, and MG63.2) have significantly higher rates of pulmonary metastasis reported in the literature with a 200-fold increase in MG63.2, 100% efficacy of DLM8 relative to no pulmonary metastases in DUNN, 100% efficacy of SAOS-LM7, and a 90% efficacy of K7M2 relative to 33% of K12. Additionally, MG63.2, DLM8, SAOS-LM7 and K7M32 cells were reported to show greater migration and invasion than their low-metastatic counterparts: MG63, DUNN, Saos2 and K12 [37, 51-53]. MG63.2 is reported to have weaker heterotypic adhesion than MG63, while K7M2 have higher initial rates of adhesion but no difference in ultimate adhesion [37, 53].

DUNN, DDLM8, K12, and K7M2 cell lines were a gift from Dr. D. Thamm (Colorado State University), MG63 and MG63.2 cell lines were a gift from Dr. D. Duval (Colorado State

University), and Saos2 and SAOS-LM7 a gift from Dr. E. S. Kleinerman (MD Anderson Cancer Center). All cell lines were maintained under typical culture conditions at 37°C and 5% carbon dioxide concentration in Dulbecco's Modified Eagle Medium (DMEM) (Sigma). DMEM was supplemented with 10% fetal bovine serum (Atlas Biologicals), 20mM Hepes (Sigma), and 100 Units/ml penicillin with 100 µg/ml streptomycin (Fisher Scientific-Hyclone). Cell lines were not independently authenticated or tested for contamination by us.

5.5.2 Immunofluorescence microscopy

Cells were cultured on indicated substrate for 48 hours. Cells were stained with Wheat Germ Agglutinin, Alexa Fluor 594 Conjugate (Molecular Probes). Cells were fixed in 4% paraformaldehyde then stained with AlexaFluor 488 Phalloidin and DAPI (Molecular Probes). Cells were imaged under a 20X objective on a Zeiss Axioplan 2 fluorescence microscope (Zeiss, Thornwood, NY, USA) using filter sets: DAPI BP 445/50 blue filter, HQ Texas Red BP 560/40, and Green BP 474/28.

5.5.3 Preparation of surfaces

Three different surfaces were prepared for this work from either a #1.5 22mm x 22mm glass coverslip (Richard Allen Scientific) or a #2 22mm x22mm siliconized glass coverslip (Hampton Research). The formulated surfaces follow: Glass Detergent washed and Air dried (GDA), Glass Acid etched and Air dried (GAA), and Silconized Ethanol Treated (SET). GAA and GDA surfaces were initially prepared by sonication for 30 min. in a mild detergent solution. Following sonication, the coverslips were sequentially rinsed with milliq (MQ) water, isopropyl alcohol (IPA), and a second rinse with MQ water prior to any further downstream processing. In the case of the GDA surface, no further downstream processing was required and the surfaces were blown dry with sterile 0.2 µm filtered nitrogen with an air gun from an in house boil off nitrogen source. GAA surface was subjected to a downstream 1M hydrochloric acid etching at 60°C for 12-16 hours. After the etching period, the coverslips went through the same rinse process described above (MQ to IPA back to MQ) before being blown dry in the same manner as the GDA surface. The SET surface was subjected to a rinse in 100% ethyl alcohol and then they were blown dry with the nitrogen gun to ensure removal of any residual liquid and debris. Prior to use in cell culture, all surfaces were exposed to UV sterilization to minimize potential contamination risks.

5.5.4 Contact angle measurements

Contact angles for different substrates were measured using sessile drop method by Rame Hart Goniometer (Model # 100_25_M). 3 microliter of miliQ water were placed on XYZ plane using needle. Image were captured and analyzed using Rame Hart DROP Image Advanced software. Contact angle were measured for 3 different spot on one slide and this was repeated 3 times on different slides to see variability of slide's contact angle. Representative images are shown in Appendix III: Supplementary Fig. 5.1.

5.5.5 Cell volume measurement using Scepter Cell Counter:

Volume measurements were made by the Scepter™ Handheld Automated Cell Counter, Millipore, with a 60 μm sensor. First, cells were plated in a culture dish. Once they were ready to be split, they were trypsinized and re-suspended in 1-X PBS. After checking that cell density was in the operating range (10,000–500,000 cells/mL) the Scepter sensor was submerged in the cell suspension. The upper and lower gates were adjusted to remove debris information, and cell volume information recorded.

The distributions of cell volumes was well approximated by a log-normal distribution. Thus we log-transformed and calculated the mean and the standard error of the mean of the resulting normal distribution. The mean cell volume for the cell lines were 1.01, 1.12, 1.60, 1.41, 1.49, 1.43, 1.66, 1.80 picoliter for DUNN, DLM8, K12, K7M2, SAOS2, LM7, MG63 and MG63.2, respectively. Thus the percentage differences between the volume of the high-met line from the low-met line are -10.9% (D lines), +11.9% (K lines), +4% (S lines) and -8.4% (M lines). We performed t-tests against the null hypothesis that the volume data for both pairs of each paired line came from a distribution with the same mean. The mean volumes for the two partners in a paired line were statistically different from each other, with p-values much smaller than 0.05 in each case.

5.5.6 Image Processing

Images of isolated cells that were not in contact with other cells were chosen. To ensure adequate statistical power we picked a sample of 100 such cell images (so that the power of the test for comparing means would be 80% at 1% significance level for a half-standard deviation effect size). Images were collected blind in the sense that the students doing the imaging were not previously aware of any differentiating characteristics discovered. The images were collected at one time for each cell line on each surface. The image processing involved three distinct steps; enhancement, conversion into binary format, and automated cropping of each cell for measurement of shape metrics. Three channels were captured as described above. Prior to processing the images were converted into 16 bit TIFF images. The exported TIFF images were loaded into MATLAB where the actin, membrane, and nuclei channels were enhanced separately by contrast stretching. The actin and membrane images were combined into a single TIFF images to get full characterization of the shape. Finally, erosion with a three pixel mask was applied to sharpen edge boundaries. The enhanced TIFF images of the combined membrane-actin image and separate nuclei image were exported into ImageJ analysis for manual conversion into binary formatted images by thresholding (a representative example is shown in Appendix III: Supplementary Fig. 5.2). Once the images had been converted into binary format, they were again loaded into

MATLAB for shape analysis. Segmentation was achieved through use of the built-in MATLAB function toolbox so each cell could be individually cropped and reconstructed in a new image in which shape measurements (listed in Supplementary Table 1) could be made on both the cell and corresponding nuclei and stored for statistical analysis. Minimum Bounding Circle was found using MATLAB function minboundcircle, open source code developed by John D'Errico [54]. Scripts for image processing will be made available upon request.

5.5.7 Data Analysis

t-Test

Individual cell metrics were compared as discussed in the results section utilizing the built in MATLAB ttest2 function which returns a test for the null hypothesis that the data come from independent random samples with normal distributions and equal means without assuming equal variance. This is a two-tailed test. The null hypothesis is initially rejected at a 5% significance level. All 29 parameters are then retested with the significance level determined by the Holm-Bonferroni correction for multiple tests.

Principal Component Analysis (PCA):

PCA is a method to project each sample in specific dimension to a space with equal or smaller dimension. This process is done in such a way that the first principal component has the maximum variance, second principal component has the next maximum variance, and this rule continues for subsequent components. The principal component vectors also form an orthogonal basis. We used singular value decomposition (SVD) to perform PCA on the data. First, data was standardized so that mean of new data is zero and standard deviation is 1. Then, SVD of the data was computed and the principal components extracted from the right singular vectors of the data.

Each data point was then projected into the space formed by the first three principal components, and was plotted for visualization. The variance captured by the first three principal components lie in the range 44%-47% of the total variance for every comparison made.

Nonmetric Multidimensional Scaling (NMDS)

We performed separate analyses for each of the three surfaces (GAA, GDA, SET) and each paired cell line (D, K, S and M). First, each of the 29 cell morphology variables were relativized by dividing each value by the maximum value. Statistical software R (version 3.1.2) and package vegan were used to perform all statistical analyses. The Bray-Curtis dissimilarity index was used to perform NMDS. Based on observed stress, convergence behavior, Shepard plots, and parsimony, k was chosen to be 3.

The ordination pattern was scaled as follows before plotting. First, centering was done to move the origin to the average of the axes. Second, principal components were used to rotate the configuration so that the variance of points was maximized for the first dimension, with the second dimension explaining the maximum variance of points unexplained by the first. We then displayed the ordination pattern in 2-space. For the factor "Metastatic capacity" (with levels Low and High) and the factor "cell line pair" (with levels D, K, M, and S), we generated two separate color-coded plots with 95% confidence ellipses and labeled locations of the level centroids. For GAA, GDA, and SET, observed stresses were approximately 0.07, 0.07, and 0.08.

Permutational multivariate analysis of variance using Bray-Curtis distance between cells (PERMANOVA) was used to obtain R^2 values for "metastatic capacity". Specifically, R^2 is a measure of the proportion of the data (distance) variation explained by "metastatic capacity".

Machine Learning

A multilayer perceptron (MLP) neural network with one hidden layer, adapted with permission from a version used by Dr. Charles Anderson for teaching [55], was used to classify data, and is available from the corresponding author upon request. A back-propagation learning algorithm, which uses a Scaled Conjugate Gradient, SCG, was used to design the MLP and tanh(x) was used as the activation function. The SCG was adapted from Nabney's netlab library [56, 57]. Each data set was partitioned into test, training and validation data at 50%, 25%, and 25% of data respectively. The test and training data sets were used to find the best attribute combinations, number of hidden units and weight parameter values in the non-linear logistic regression model. Initial parameters are chosen randomly. Training data was used to fit parameters by maximizing a likelihood function; testing data was then used to calculate the percentage of cells classified correctly (test percent). To optimize the model, training and test data were repartitioned and an average test percent was calculated for different attribute combinations and function structure; we selected the optimal attribute combination and function structure based on the maximum average test percent. After the function structure was chosen the test and training data sets were combined for one last round of optimization of the weights. The optimized model was then used to predict the class that each individual cell in the validation data belongs to with no further adjustment of parameters.

To test the accuracy on random samples of cells from each population, after identifying the function structure with the training and test data, we took 100 random paired samples of 10 cells each from the validation data set. The percentage of cells in each sample predicted to be class 1 are recorded (*P*). Thus the percentage of cells predicted to be class 2 = 1- *P*. A decision threshold was determined utilizing the false negative rate (FNR) and true positive rate (TPR). When *P* was bigger than the decision threshold, the sample was classified as class 1, and when it was smaller than the decision threshold as class 2. As detailed in Supplementary Table 4**,** the threshold was optimal at 0.6. From the total 100 pairs, the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) were calculated. Using this information, accuracy, false negative rate (FNR) and true negative rate (TPR) were calculated as defined below:

$$Accuracy = \frac{TP + TN}{TN + FN + TP + FP}$$

$$FNR = \frac{FN}{FN + TP}$$

$$TPR = \frac{TP}{FN + TP}$$

5.6 Acknowledgements

We gratefully acknowledge Dr. Jennifer DeLuca and Dr Diego Krapf for helping us start these experiments, and Dr. Chuck Anderson for permission to use and adapt his neural network algorithm. We thank Dr. Dawn Duval and Dr. E. S. Kleinerman for the gift of cell lines. This work was supported through an NSF CAREER grant (PHY-1151454 to A.P.).

5.7 Author Contributions

A.P., S.M.L. and D.T. designed experiments, S.M.L., E.A., J.C., B.S. performed experiments, S.M.L., E.A., J.M. and A.P. analyzed data, J.M. wrote image processing code, K.S. and E.A. performed the machine-learning algorithm based analysis including code writing and results analysis, P.T. performed the NMDS analysis and A.P. and SML wrote the paper. All authors participated in editing, and read and approved the final manuscript.

REFERENCES

[1] Siegel, R., et al., *Cancer statistics, 2014.* CA Cancer J Clin, 2014. **64**(1): p. 9-29.

[2] Gupta, G.P. and J. Massague, *Cancer metastasis: building a framework.* Cell, 2006. **127**(4): p. 679-95.

[3] Suresh, S., *Biomechanics and biophysics of cancer cells.* Acta Biomater, 2007. **3**(4): p. 413-38.

[4] Makale, M., *Cellular mechanobiology and cancer metastasis.* Birth Defects Res C Embryo Today, 2007. **81**(4): p. 329-43.

[5] Chambers, A.F., A.C. Groom, and I.C. MacDonald, *Dissemination and growth of cancer cells in metastatic sites.* Nat Rev Cancer, 2002. **2**(8): p. 563-72.

[6] Xu, W., et al., *Cell stiffness is a biomarker of the metastatic potential of ovarian cancer cells.* PLoS One, 2012. **7**(10): p. e46609.

[7] Li, Q.S., et al., *AFM indentation study of breast cancer cells.* Biochem Biophys Res Commun, 2008. **374**(4): p. 609-13.

[8] Cross, S.E., et al., *AFM-based analysis of human metastatic cancer cells.* Nanotechnology, 2008. **19**(38): p. 384003.

[9] Cross, S.E., et al., *Nanomechanical analysis of cells from cancer patients.* Nat Nanotechnol, 2007. **2**(12): p. 780-3.

[10] Guck, J., et al., *Optical deformability as an inherent cell marker for testing malignant transformation and metastatic competence.* Biophys J, 2005. **88**(5): p. 3689-98.

[11] Guck, J., et al., *The optical stretcher: a novel laser tool to micromanipulate cells.* Biophys J, 2001. **81**(2): p. 767-84.

[12] Runge, J., et al., *Evaluation of single-cell biomechanics as potential marker for oral squamous cell carcinomas: a pilot study.* Oral Dis, 2014. **20**(3): p. e120-7.

[13] Berx, G. and F. van Roy, *Involvement of members of the cadherin superfamily in cancer.* Cold Spring Harb Perspect Biol, 2009. **1**(6): p. a003129.

[14] Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

[15] Cavallaro, U. and G. Christofori, *Cell adhesion and signalling by cadherins and Ig-CAMs in cancer.* Nat Rev Cancer, 2004. **4**(2): p. 118-32.

[16] Chen, J.C., Y.C. Fong, and C.H. Tang, *Novel strategies for the treatment of chondrosarcomas: targeting integrins.* Biomed Res Int, 2013. **2013**: p. 396839.

[17] Rathinam, R. and S.K. Alahari, *Important role of integrins in the cancer biology.* Cancer Metastasis Rev, 2010. **29**(1): p. 223-37.

[18] Tullberg, K.F. and M.M. Burger, *Selection of B16 melanoma cells with increased metastatic potential and low intercellular cohesion using Nuclepore filters.* Invasion Metastasis, 1985. **5**(1): p. 1-15.

[19] Paszek, M.J., et al., *Tensional homeostasis and the malignant phenotype.* Cancer Cell, 2005. **8**(3): p. 241-54.

[20] Kenny, P.A., et al., *The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression.* Mol Oncol, 2007. **1**(1): p. 84-96.

[21] McGrail, D.J., et al., *Actomyosin tension as a determinant of metastatic cancer mechanical tropism.* Phys Biol, 2015. **12**(2): p. 026001.

[22] Bakal, C., et al., *Quantitative morphological signatures define local signaling networks regulating cell morphology.* Science, 2007. **316**(5832): p. 1753-6.

[23] Straw, R.C., et al., *Canine mandibular osteosarcoma: 51 cases (1980-1992).* J Am Anim Hosp Assoc, 1996. **32**(3): p. 257-62.

[24] Kirpensteijn, J., et al., *Prognostic significance of a new histologic grading system for canine osteosarcoma.* Vet Pathol, 2002. **39**(2): p. 240-6.

[25] Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* J Clin Invest, 2009. **119**(6): p. 1420-8.

[26] Odenwald, M.A., J.R. Prosperi, and K.H. Goss, *APC/beta-catenin-rich complexes at membrane protrusions regulate mammary tumor cell migration and mesenchymal morphology.* BMC Cancer, 2013. **13**: p. 12.

[27] Cowden Dahl, K.D., et al., *The epidermal growth factor receptor responsive miR-125a represses mesenchymal morphology in ovarian cancer cells.* Neoplasia, 2009. **11**(11): p. 1208-15.

[28] Loukopoulos, P. and W.F. Robinson, *Clinicopathological relevance of tumour grading in canine osteosarcoma.* J Comp Pathol, 2007. **136**(1): p. 65-73.

[29] Morello, E., M. Martano, and P. Buracco, *Biology, diagnosis and treatment of canine appendicular osteosarcoma: similarities and differences with human osteosarcoma.* Vet J, 2011. **189**(3): p. 268-77.

[30] Ottaviani, G. and N. Jaffe, *The epidemiology of osteosarcoma.* Cancer Treat Res, 2009. **152**: p. 3-13.

[31] Link, M.P., et al., *The effect of adjuvant chemotherapy on relapse-free survival in patients with osteosarcoma of the extremity.* N Engl J Med, 1986. **314**(25): p. 1600-6.

[32] Ward, W.G., et al., *Pulmonary metastases of stage IIB extremity osteosarcoma and subsequent pulmonary metastases.* J Clin Oncol, 1994. **12**(9): p. 1849-58.

[33] Kaste, S.C., et al., *Metastases detected at the time of diagnosis of primary pediatric extremity osteosarcoma at diagnosis: imaging features.* Cancer, 1999. **86**(8): p. 1602-8.

[34] Yonemoto, T., et al., *Prognosis of osteosarcoma with pulmonary metastases at initial presentation is not dismal.* Clin Orthop Relat Res, 1998(349): p. 194-9.

[35] Jaffe, N., et al., *Single and multiple metachronous osteosarcoma tumors after therapy.* Cancer, 2003. **98**(11): p. 2457-66.

[36] Kager, L., et al., *Primary metastatic osteosarcoma: presentation and outcome of patients treated on neoadjuvant Cooperative Osteosarcoma Study Group protocols.* J Clin Oncol, 2003. **21**(10): p. 2011-8.

[37] Su, Y., et al., *Establishment and characterization of a new highly metastatic human osteosarcoma cell line.* Clin Exp Metastasis, 2009. **26**(7): p. 599-610.

[38] Cavallaro, U. and G. Christofori, *Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough.* Biochim Biophys Acta, 2001. **1552**(1): p. 39-45.

[39] Grinnell, F. and M.K. Feld, *Fibronectin adsorption on hydrophilic and hydrophobic surfaces detected by antibody binding and analyzed during cell adhesion in serum-containing medium.* J Biol Chem, 1982. **257**(9): p. 4888-93.

[40] Pincus, Z. and J.A. Theriot, *Comparison of quantitative methods for cell-shape analysis.* J Microsc, 2007. **227**(Pt 2): p. 140-56.

[41] Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

[42] Anderson, M.J., *A new method for non-parametric multivariate analysis of variance.* Austral Ecology, 2001. **26**: p. 32-46.

[43] Treiser, M.D., et al., *Cytoskeleton-based forecasting of stem cell lineage fates.* Proc Natl Acad Sci U S A, 2010. **107**(2): p. 610-5.

[44] Wolf, K., et al., *Compensation mechanism in tumor cell migration: mesenchymal-amoeboid transition after blocking of pericellular proteolysis.* J Cell Biol, 2003. **160**(2): p. 267-77.

[45] Sanz-Moreno, V. and C.J. Marshall, *The plasticity of cytoskeletal dynamics underlying neoplastic cell migration.* Curr Opin Cell Biol, 2010. **22**(5): p. 690-6.

[46] Sahai, E. and C.J. Marshall, *Differing modes of tumour cell invasion have distinct requirements for Rho/ROCK signalling and extracellular proteolysis.* Nat Cell Biol, 2003. **5**(8): p. 711-9.

[47] Liu, Y.J., et al., *Confinement and low adhesion induce fast amoeboid migration of slow mesenchymal cells.* Cell, 2015. **160**(4): p. 659-72.

[48] Eefting, D., et al., *Assessment of interobserver variability and histologic parameters to improve reliability in classification and grading of central cartilaginous tumors.* Am J Surg Pathol, 2009. **33**(1): p. 50-7.

[49] Engers, R., *Reproducibility and reliability of tumor grading in urological neoplasms.* World J Urol, 2007. **25**(6): p. 595-605.

[50] Yin, Z., et al., *A screen for morphological complexity identifies regulators of switch-like transitions between discrete cell shapes.* Nat Cell Biol, 2013. **15**(7): p. 860-71.

[51] Asai, T., et al., *Establishment and characterization of a murine osteosarcoma cell line (LM8) with high metastatic potential to the lung.* Int J Cancer, 1998. **76**(3): p. 418-22.

[52] Jia, S.F., L.L. Worth, and E.S. Kleinerman, *A nude mouse model of human osteosarcoma lung metastases for evaluating new therapeutic strategies.* Clin Exp Metastasis, 1999. **17**(6): p. 501-6.

[53] Khanna, C., et al., *An orthotopic model of murine osteosarcoma with clonally related variants differing in pulmonary metastatic potential.* Clin Exp Metastasis, 2000. **18**(3): p. 261-71.

[54] D'Errico, J. *minbouncircle.m*. 2015 [cited 2015 July 3]; Available from: http://www.mathworks.com/matlabcentral/fileexchange/34767-a-suite-of-minimal-bounding-objects/content/MinBoundSuite/minboundcircle.m.

[55] Anderson, C. *CS545: Machine Learning*. 2015 [cited 2015 June 29, 2015]; Available from: http://www.cs.colostate.edu/~anderson/cs545/index.html/doku.php.

[56] Møller, M.F., *A scaled conjugate gradient algorithm for fast supervised learning.* Neural Networks, 1993. **6**(4): p. 525-533.

[57] Nabney, I.T., *Netlab: Algorithms for Pattern Recognition*. 2002, London: Springer.

CHAPTER 6: CONCLUSION

6.1 Mathematical and experimental studies of cellular decisions

This thesis has explored some aspects of cellular decisions, a phrase used here to describe cell state changes under external or internal signals, using mathematical modeling as well as quantitative data analysis of experiments. While the range of topics studied is broad, encompassing information transfer in signal transduction, modularity in biological switches and identification of cell state changes using shape changes, the work is united by the common theme of mathematical and quantitative tools utilized. In each case, this work has opened new avenues of research and investigation. This chapter concludes with a brief overview of some of the most interesting questions that have arisen as a result of this research and discusses avenues for future work.

6.2 Conclusions from Chapter 3: Cross-talk and information transfer in mammalian and bacterial signaling

In Chapter 3, we utilized information theory to study the transmission of information in simple signaling networks based upon the SMAD signaling pathway of the TGF-β family and the two-component signaling systems of bacteria. While signaling is often thought of as communicating a binary state, e.g. 'on' or 'off' via activation (or repression) of gene transcription, cells can in theory transmit much more information than 1 bit, which could be used to perform more complex information processing than a simple binary decision. While it is not known if signal transduction networks do indeed utilize all the information available to them, some sensory cells have highly evolved sensing opportunities and one could argue that evolution should have optimized information transmission of signal transduction networks similarly.

We found that for high levels of information transmission in a system where many ligands signal through a small set of common receptors and signaling proteins such as the TGF- pathway modeled, multiple signaling proteins are needed. This provides an information processing based explanation for the multiplicity of signaling proteins involved in signal transduction networks. Production of any given protein is costly to a cell and one may assume that this multiplicity must be as a result of some evolutionary advantage; information fidelity and processing may well be the driving force for this.

We also observed that for systems with large size and phosphorylation relay, information transfer is quite robust in the presence of high cross-talk; as long as cross-talk was below approximately 70%, the cell was theoretically able to distinguish between external signals with high accuracy. This counter-intuitive finding provides an information theory based explanation for the acquisition of new signaling pathways. As discussed in Chapter 3, a mutation in a signaling protein which allows for binding of novel molecules but also continues to bind with prior molecules could lead to preferential binding for an existing function or acquisition of a new function. By tolerating a high level of cross-talk, these similar structures could still provide the cell an increase in information transfer without having to, upon initial mutation, provide a completely separate signaling channel. This may well explain the origins of several BMP signaling pathways, where BMP2 and BMP4 share 92% of structural similarity and demonstrate cross-talk, but have non-redundant cellular function. These two proteins may have originated by mutations of a single ancestral BMP protein. The increase in information transmission by such mutations could be a source of positive evolutionary selection.

On the other hand, we found that the bacterial two-component signaling pathway, with a smaller system size and phospho-transfer rather than relay, was highly sensitive to cross talk.

This provides a potential explanation for why there are few examples of cross-talk between the hundreds of structurally similar two-component pathways in a bacterium, and why experimentally forced cross-talk is lost over generations of cellular division. Cross-regulation, or the exploitation of cross-talk, is not possible when the interference between two pathways leads to loss of information transfer, as was observed in our model of the two-component signaling pathway.

Further work on this topic should probe the mechanisms behind the effects of system size. Further work is also needed to understand how gene transcription networks can interpret signals coming from systems with an innately high level of cross-talk. Additionally, this model should be expanded to understand what happens when there is cross-talk between more than two pathways at the same time. This is particularly relevant for TGF-β signaling and BMP signaling, since both have at least three SMAD homolog's that are involved in information transfer from the receptor to the nucleus. Finally, our analysis also leads to the design of experiments to be performed that can confirm or falsify our predictions and uncover how cells make sense of the world in the presence of cross-talk. For example, it would be interesting to force cross-talk between two-component signaling pathways in bacterium with small and with large system size, to observe the tolerance of the bacteria to this cross talk. Our work would prompt the hypothesis that in a small system size, the cross talk would decrease over generations whereas in the larger system size it may be more tolerated or perhaps even exploited.

6.3 Conclusions from Chapter 4: Loads bias genetic and signaling switches in synthetic and natural systems

For a genetic toggle switch with two mutually repressing proteins we demonstrated that while the connection with a down-stream load through a binding partner does not alter steady state properties of the switch, it can drastically change the dynamic properties. To better understand how this occurs, we utilized a novel potential landscape analysis to show that the binding partner skews the underlying quasi-potential, making one state significantly more stable than the other. In practice, a genetic toggle switch that is significantly skewed towards one side may never properly function as a switch. This result underscores the importance of factoring for downstream loads and the consequences of their existence when designing connected synthetic circuits.

This finding also provides a strategy for making artificial switches tunable. By adding a load to one side of the toggle, one can bias the switch toward the opposite side, which creates a switch that is designed to only turn on with special limited circumstances. The addition of a load on both sides can deepen the quasi-potential landscape which stabilizes the switch, a useful feature for synthetic components in low concentrations that are subject to noise.

This work also provides an explanation for naturally observed complexities on toggle switches - some systems demonstrate a positive feedback motif that stabilizes the side it is on. This positive feedback may ensure that the switch operates despite intracellular variability. A topic of future work would be identifying switches and characterizing any loads on the system; identification of positive feedback is a potential target for evolutionary refinement of the switch. Commonly, the additional complexities of natural protein signaling networks are omitted in analysis as simple models often qualitatively capture the networks behavior, however these additional protein interactions, such as the presence of loads and feedback, may serve a function in tuning and stabilizing genetic switches. The production of any protein is costly to a cell and from an evolutionary perspective some argue that there must be a benefit to the existence of each protein and interaction; the quasi-potential landscape explored in this work and the tuning and stabilization properties of additional protein interactions in naturally occurring switches may explain why those loads and feedback exist. Future work could explore the effects of the addition and inhibition of loads and feedback in naturally occurring cellular systems to probe this question further.

Our analysis underscores the importance of incorporating loads when simulating models of switches in natural and synthetic systems. Mathematical analysis of switch-like motifs therefore would do well to at least include a load on their output proteins, to incorporate the possible effects of load induced modulation on the circuit. As the field of synthetic biology progresses, and the cycle of design, build, test becomes increasingly more attainable rather than brute force methods, it would be interesting to note if more accurate builds and more success is achieved when loads are accounted for in the design phase.

6.4 Conclusions from Chapter 5: Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas

In Chapter 5, advanced statistical analysis and machine learning methods were combined with *in vitro* experiments and sophisticated image processing to create a toolbox to assess spread cell shape of cancer cells. This work studied isogenic paired lines of highly metastatic osteosarcomas derived from less metastatic parental cells. This work contributes to our understanding of metastasis in two ways: the first, our findings on cellular shape changes can aid mechanistic knowledge of metastasis and provide an opportunity to interpret genetic changes in metastasis through understanding those effects on spread cell shape; the second is that we have developed an initial toolbox which with refinement could potentially be applied to clinical settings for prognostication and classification.

6.4.1 Contributions to understanding of cellular shape in metastasis

The concept of cellular shape as an insight into cancer aggressiveness is not novel, however the use of spread cell shape on various adhesive surfaces to probe the mechanisms of mechanical property changes is less explored.

Analysis of shape showed that three of the four paired lines had a similar class of shape changes characterized by an elongated spindly shape with higher variability. The fourth paired line showed a different class of shape changes with metastatic cells having a rounder shape. In both cases, the distribution of shape characteristics showed greater variability for the highly metastatic cell line which corresponds with histopathology markers of aggressiveness - anisocytosis and anisokaryosis meaning variability in cell and nuclear shape/size.

The genetic changes that drive acquisition of invasive properties affect shape change in various ways. For example, changes necessary for detachment of a tumor could be due to many different processes such as down-regulation of adhesive proteins, changes in cellular contraction due to Rho-ROCK signaling, or softening of the cell due to changes in vimentin and keratin; each of these changes would affect cell shape in potentially different ways which could be assessed with less expensive image analysis rather than detailed protein assays. This would require characterization of cell shape changes based upon inhibition or amplification of the various proteins in future experiments paired with image processing. Many studies regarding up- or down-regulation of proteins involved in metastatic disease have qualitative cellular shape changes described in passing, but there is a missed opportunity for quantitative analysis of shape changes. The toolbox presented in Chapter 5 provides one strategy for extracting more information from cellular experiments and providing better understanding of the relationship between genetic changes, changes in mechanical properties, and the resulting changes in cellular shape in cancer. Identifying and understanding the full typology of shape changes could have a major impact on our knowledge of metastasis and its relation to the cellular cytoskeleton. It may be eventually possible to read out genetic changes corresponding to specific changes in cell shape but much groundwork is needed to link genetic changes with specific shape.

6.4.2 A toolbox for automated cell shape analysis and metastatic prognostication

This work developed an initial toolbox that works to extract large amounts of information from cellular experiments through quantifiable shape metrics. With refinement, this concept could provide personalized high-throughput data on cancer cell shape, and ultimately could be used to guide clinical decision making in conjunction with the gold standard of stage and grade.

There was a significant overlap between shape parameter distribution for low vs highly metastatic lines, however advanced statistical analysis including principle component analysis (PCA) and non-metric multidimensional scaling (NMDS) was able to show that the cell shape parameters clustered differently under some culture conditions. These differences were not significant enough to draw boundaries in multidimensional non-linear space to separate cell types, but they were sufficient to utilize neural network machine learning to classify cells with high accuracy.

Further work is needed to validate and refine this framework with use of other cell lines, primary tumor lines and other cancer types as well as other species. As discussed in Chapter 1, clinicians utilize multiple pieces of data to formulate a clinical treatment plan and prognosis including cancer stage and grade, sometimes even multiple grades from different grading schemes as in canine mast cell tumors. While some grading schemes have been developed to address or limit inter-observer variability noted in histopathology grading, there is still a high level of variability in some cancer types. Extension of this work provides an additional tool for clinicians if we are able to evolve this framework to provide accurate predictions.

An ideal application would be applying a high through-put, automated imaging analysis of cancer cells - either cultured *in vitro* as our work was or perhaps even tissue biopsy samples - to analyze hundreds of cancer cells, extract cell shape parameters, and through a highly trained neural network, output a mathematical probability of metastasis. If subtypes of cancer can be read out in shape, for example the type-1 vs type-2 classes of cell shape changes observed in this work, and those types are correlated to response to various treatments, this tool could guide clinical decision making. For example, if we were able to analyze banked tissue samples of tumors and pair the classification of shape with known outcomes and responses to various therapies, we could utilize machine learning to predict a new sample's response to different therapies.

This type of personalized, predictive medicine is not novel; there are many cancer assays which seek to predict which cancers are likely to respond to specific treatments, mostly based upon mRNA, immunohistochemistry and protein biomarkers. Cellular shape has yet to be used in this way and provides an opportunity for potentially less expensive, faster results as cellular shape analysis in histopathology is already utilized in nearly every cancer diagnosis. While this application would require extensive further refinement of imaging algorithms, access to and analysis of banked tissue samples, and advanced training neural networks, this concept is achievable. It represents a very useful, exciting application from pairing experiment with mathematical modeling and analysis of the supremely important cellular 'decision' of metastasis.

6.5 Concluding remarks

This work demonstrates the successful integration of biology and mathematical modeling to assess cellular decision making in mammalian, bacterial and synthetic systems, combining a variety of techniques, including several novel applications, to further understanding of these systems. There is great opportunity for future work on each of the systems studied, which could advance basic science understanding as well as provide clinically useful applications.

APPENDIX I: SUPPLEMENTARY INFORMATION FOR CHAPTER 3

7.1 Detailed Methods

7.1.1 Model Development: Adaptation to Stochastic Modeling of Smad Pathway

The Smad signaling cascade for both TGF-β and BMP pathways has been mathematically modeled by a number of research groups [5, 7]. These groups used a deterministic method to simulate the signaling, however a stochastic method is necessary for information theory. Using the experimentally-derived kinetic parameters from Schmierer et al., and model components utilized by Nakabayashi and Sasaki, the stochastic model was developed. Schmierer et al. used an estimated nuclear volume of 1 pl and a cytoplasmic volume of 2.3 pl. These volumes were utilized to convert the necessary kinetic parameters from Nakabayashi and Sasaki from molar units to molecular units. A sample conversion from molar to molecular units is shown below in Equations S1-S3. Additionally, after preliminary investigation revealed that nuclear shuttling did not affect information transmission, this component was not included in the final model. The species used, initial values, reactions, and parameters are summarized in Table S3.1-2.

Table S3.1 Smad Model Reactions

	Reaction	Forward Rate	Reverse Rate	Description
1	R1+L1 <-> R1:L1	k_X	δ_X	Ligand binds receptor
2	R2+L2 <-> R2:L2	k_Y	δ_Y	Ligand binds receptor
3	R1 -> null	δ_{R1}		Receptor internalization/degredation
4	R2 -> null	δ_{R2}		Receptor internalization/degredation
5[2]	ri+Rj:Lj<->ri:Rj:Lj	γ_{aij}	γ_{bij}	rsmad binds receptor/ligand
6[2]	ri:Rj:Lj->Rj:Lj + ri:p	γ_{cij}		receptor phosphorylates and releases rsmad
7	r:p-> r	δ_P		rsmad-p dephosphorylation
8	r:p + C <-> r:p:C	μ	λ	rsmad-p binds cosmad

Note: i = 1 or 2 for rsmad1 or rsmad2; j = 1 or 2 for Receptor/Ligand 1 or 2. All reactions were taken from (Nakabayashi & Sasaki, 2009) except where noted (2), and Michaelis-Menten kinetics were used for phosphorylation of an rsmad by a receptor:ligand complex.

Table S3.2 Parameters and Initial Amounts for Smad Model

Parameter	Standard Rate	Units	Description
$k_X = k_Y$[2]	1.0e-5	1/(molecule*second)	Ligand association rate
$\delta_X = \delta_Y$[1]	5e-5	1/second	Ligand dissociation rate
$\delta_{R1} = \delta_{R2}$[1]	5e-4	1/second	Receptor degredation rate
δ_P[1]	6.6e-3	1/second	rsmad dephosphorylation rate
μ[1]	1.3e-6	1/(molecule*second)	rsmad:Cosmad association Rate
λ[1]	0.016	1/second	rsmad:Co dissociation rate
$\gamma_{a11}= \gamma_{a22}$[2]	4.0e-6	1/(molecule*second)	like R:L + rsmad association rate
$\gamma_{a12}= \gamma_{a21}$[2]	0 - 4.0e-6 (variable)	1/(molecule*second)	unlike R:L + rsmad association rate
$\gamma_{b11}= \gamma_{b22} = \gamma_{b12}= \gamma_{b21}$[2]	1.0e-4	1/second	R:L + rsmad dissociation rate
$\gamma_{c11}= \gamma_{c22}= \gamma_{c12}= \gamma_{c21}$[2]	2	1/second	Phosphotransfer/dissociation rate
L1, L2	0 – 260[1]	molecule	Ligand 1, 2
R1, R2	250[1]	molecule	Receptor 1, 2
r, r1, r2	10000 each[2]	molecule	Rsmad 1, 2
C	10000[2]	molecule	Co-Smad
r:p, r1:p, r2:p	0[1]	molecule	Phosphorylated Rsmad

Note: Parameters and values were taken from 1. (Nakabayashi & Sasaki, 2009) or 2. based on appropriate ranges from similar rates from (Nakabayashi & Sasaki, 2009).

$$\mu = 0.0018\, nM^{-1}s^{-1} = \frac{0.0018}{nM\ s} \cdot \frac{1}{2.3pL} \cdot \frac{1\ L}{1E^{-9}\ mol} \cdot \frac{1E^{12}pL}{1\ L} \cdot \frac{1\ mol}{6.022E^{23}\ molecules} = \frac{1.3\ E^{-6}}{molecules\ s}$$ **Equation S1**

$$[R] = 1\ nM \cdot \frac{1}{1\ pL} \cdot \frac{1}{2.3pL} \cdot \frac{1E^{-9}mol}{1\ L} \cdot \frac{1\ L}{1E^{12}pL} \cdot \frac{6.022E^{23}molecules}{1\ mol} = 261\ Receptors$$ **Equation S2**

$$\delta_p = 6.6E^{-3}\ s^{-1}$$ **Equation S3**

Additionally, for ease of calculation, 250 receptors were chosen for each type. The amount of ligand was varied from 0 to 250 molecules because, as can be seen in Figure S3.1, the maximum ligand results in a plateau of maximum RC accumulation and thus incorporates the full dynamic range.

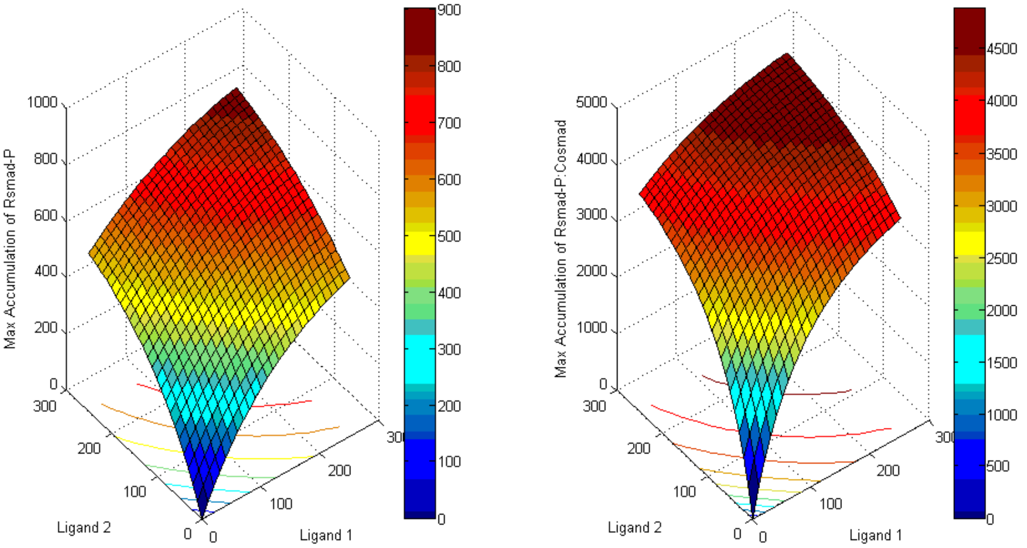


Figure S3.1. Standard Output for Two Output R and RC.
The x- and y-axes correspond to the initial ligand amount of ligand X and Y respectively. The z-axis is the average maximum accumulation of the outputs: phosphorylated RSmad and the RSmad:Co-Smad heterodimer. Note that the outputs saturate at maximum initial ligand amounts.

7.1.2 Stochastic Modeling of the Two-Component System

Bacterial two-component systems have similarly been mathematically modeled by a number of groups. We chose to base our model on one developed by Ref [1]. The same process as described above was used to adapt a deterministic model into a stochastic one. The resulting reactions and parameters are tabulated in Table S3.3-5. The dynamic range can be seen in Figure S3.2.

Table S3.3. Two-Component Model Reactions

	Reaction	**Forward Rate**	**Reverse Rate**	**Description**
1	HK1+L1 <-> HK1:L1	k_X	δ_X	Ligand binds receptor
2	HK2+L2 <-> HK2:L2	k_Y	δ_Y	Ligand binds receptor
3	HK -> HKp	K_{ap}		HK autophosphorylation
4	HK:L ->HKp:L	K_{lp}		HK autophosphorylation with ligand bound
5	HKp ->HK	K_{ad}		HK dephosphorylation regardless of ligand
6	HKi+RRj<->HKi:RRj	K_{bij}	k_d	HK binds RR regardless of phosphorylation state
7	HKp:RR ->HK:RRp	k_{pt}		HK phosphotransfer to RR
8	HK:RRp -> HK:RR	k_{pd}		HK phosphatase activity on RR
9	RRp -> RR	d_{phos}		RR dephosphorylates
10	L -> null	δ_L		Ligand degrades

Table S3.4 Parameters and Initial Values for Two-Component Model

Parameter	**Standard Rate**	**Units**	**Description**
$k_X = k_Y$[4]	3.4E-4	1/(molecule*second)	ligand association rate
$\delta_X = \delta_Y$[4]	0.5	1/second	ligand dissociation rate
k_{ap}[3]	0.001	1/second	HK autophosphorylation
k_{lp}[3]	0.1	1/second	HK autophosphorylation w/ L
k_{ad}[3]	5E-4	1/second	HK dephosphorylation
$kb_{11}=kb_{22}$[3]	3.4E-4	1/(molecule*second)	cognate RR+HK binding
$kb_{12}= kb_{21}$[4]	0 -3.4E-4 (variable)	1/(molecule*second)	non-cognate RR+HK binding
k_d[3]	0.5	1/second	RR+HK unbinding
k_{pt}[3]	1.5	1/second	phosphotransfer
k_{pd}[3]	0.05	1/second	HK phosphatase
dphos[3]	6.6E-3	1/second	RRp dephosphorylation
δ_L[4]	5E-5	1/second	ligand degradation
L1, L2[4]	0 – 250[3]	molecule	ligand 1, 2 initial amount
HK1, HK2[3]	250[3]	molecule	HK 1, 2 initial amount
RR1, RR2[3]	8824 each[3]	molecule	RR 1, 2 initial amount

Note: Initial Values and parameters were taken from 3. (Igoshin, Alves, & Savageau, 2008) or 4. based on appropriate ranges from similar rates from (Igoshin, Alves, & Savageau, 2008).

Table S3.5. Effect of Symmetric Parameter Change

		Percent Change in Information (% of Bits)			Absolute Change in Efficiency (%)		
	Fold Change	I(X,Y;Z)	I(X;Z)	I(Y;Z)	$\frac{I(X,Y;Z)}{H(X,Y;Z)}$	$\frac{I(X;Z)}{H(X;Z)}$	$\frac{I(Y;Z)}{H(Y;Z)}$
δ_P	x10	-3.43%	-1.21%	-1.09%	-1.28%	-0.20%	-0.18%
	x0.1	-5.18%	-1.98%	-2.37%	-1.97%	-0.31%	-0.37%
δ_R	x10	-40.89%	-22.45%	-22.69%	-15.25%	-3.11%	-3.15%
	x0.1	23.14%	10.08%	10.29%	10.16%	1.66%	1.69%
	x0.05	25.61%	8.95%	8.81%	10.24%	1.29%	1.27%
	x0.01	25.72%	8.74%	8.65%	10.29%	1.26%	1.25%
	x0.001	25.85%	8.79%	8.57%	10.34%	1.27%	1.24%
	x0.0001	25.86%	8.82%	8.53%	10.34%	1.28%	1.23%
δ_X δ_Y	x10	-12.18%	-6.95%	-7.02%	-3.85%	-0.84%	-0.85%
	x0.1	4.96%	4.42%	3.75%	2.95%	0.83%	0.73%
γ_A	x10	0.85%	0.53%	0.52%	1.32%	0.26%	0.26%
	x0.1	-12.39%	-4.55%	-4.89%	-3.93%	-0.49%	-0.54%
γ_B	x10	-0.16%	-0.04%	0.22%	0.92%	0.17%	0.21%
	x0.1	1.29%	6.62%	5.02%	1.49%	1.15%	0.92%
γ_C	x10	-0.13%	0.26%	-0.17%	0.93%	0.22%	0.16%
	x0.1	-0.01%	0.38%	0.50%	0.98%	0.24%	0.25%
K_x K_y	x10,000	36.31%	3.48%	3.59%	14.49%	0.49%	0.51%
	x1,000	36.07%	4.62%	4.80%	14.39%	0.66%	0.68%
	x100	35.32%	6.21%	6.22%	14.10%	0.89%	0.89%

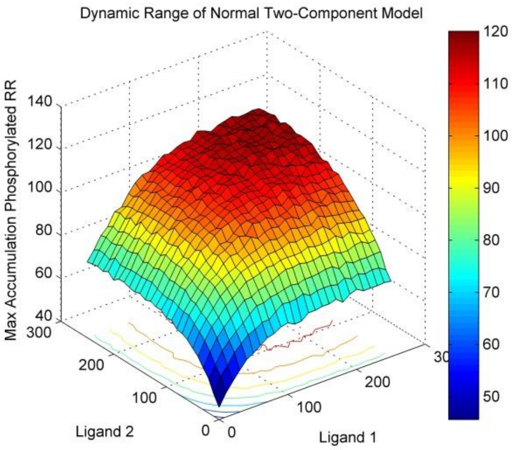


Figure S3.2. **Standard Output for Phosphorylated Response Regulator.**
The x- and y-axes correspond to the initial ligand amount of ligand X and Y respectively. The z-axis is the average maximum accumulation of the output: phosphorylated response regulator. Note that the outputs saturate at maximum initial ligand amounts. This was done with 100% cross talk and shows response regulator 1.

7.1.3 Model Simulation and Mean/Standard Deviation Calculation

The matrix of initial conditions is established as a 26 x 26 grid representing ligand values of 0 to 250 molecules in steps of ten for both ligand 1 and ligand 2, resulting in 676 possible initial conditions. For each of these initial conditions, 100 stochastic runs are simulated using SSC, a linux based stochastic simulation compiler [2]. The maximum accumulation of output was recorded for each run as averaged over a window of 50 seconds to smooth over stochastic noise. We assume that for each initial condition, the maximum accumulation of the output is normally distributed about a mean value; distributions of the data appear approximately normal. The mean and standard deviation of the maximum accumulation from the 100 runs for each initial condition are calculated and used for information calculation as described below. The output from each run through the prior is a 26 x 26 matrix of mean maximum output values, μ for each output (z or z_1 and z_2) as well as a 26 x 26 matrix of standard deviation values for each output (z or z_1 and z_2).

7.1.4 Information Calculation

The standard equations for information calculation are listed below. The two forms of mutual information are given by I(X,Y,Z) in Equation S4 and I(X,Z) in Equation S5 (Reza, 1994).

Equation S4

$$I(X,Y,Z) = \sum_{z\in Z}\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j,z)\log\left(\frac{p(x_i,y_j,z)}{p(z)q(x_i,y_j)}\right)$$

Equation S5

$$I(X,Z) = \sum_{z\in Z}\sum_{j=1}^{m} p(x_i,y_j,z)\log\left(\frac{p(x_i,y_j,z)}{p(z)q(x_i,y_j)}\right)$$

Where n = number of X inputs, m = number of Y inputs. $q(x_i,y_j)$ is the prior, or the probability of having a given set of initial ligand amounts. We assume an even prior so for all i and j, we have Equation S6.

$$q(x_i,y_j) = \frac{1}{nm}$$

Equation S6

The specific probabilities are defined by Equation S7-9.

$$p(x_i,y_j,z) = p(z|x_i,y_j)q(x_i,y_j)$$

Equation S7

Equation S8

$$p(x_i,z) = \sum_{j=1}^{m} p(z|x_i,y_j)q(x_i,y_j)$$

Equation S9

$$p(z) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(z|x_i,y_j)q(x_i,y_j)$$

For information calculation, the normal distribution of output is integrated over all possible output values by binning z. Z is divided into bins that range from zero to z_{max} with a stepsize z_{step}. For the ith bin, the interval of the integral is given by: $\left[z_i - \frac{z_{step}}{2}, z_i + \frac{z_{step}}{2}\right]$.

7.1.5 For information calculation based on one output:

- z_{max} is set as the maximum z observed across the prior plus the maximum recorded standard deviation across the prior.
- z_{step} is set at one.
- The entire range of z from 0 to z_{max} is calculated.
- The normal cdf is taken, centered at each z value, on the interval $\left[z_i - \frac{z_{step}}{2}, z_i + \frac{z_{step}}{2}\right]$, as shown in Equation S10 below [4].

$$P(z|x_i, y_j) = \int_{Z-\frac{z_{step}}{2}}^{Z+\frac{z_{step}}{2}} \frac{1}{\sqrt{(2\pi\sigma_{x,y,z}{}^2)}} \exp\left(-\frac{[Z-\mu_{x,y,z}]^2}{2\sigma_{x,y,z}{}^2}\right) dZ \qquad \textbf{Equation S10}$$

- This value is then used in Equation S7-9 to calculate the appropriate probabilities and information.

7.1.6 For information calculation based on two outputs:

- z_{max1} is set as the maximum z_1 observed across the prior plus the maximum recorded standard deviation for z_1 across the prior.
- z_{max2} is set as the maximum z_2 observed across the prior plus the maximum recorded standard deviation for z_2 across the prior.
- z_{step} is set at ten.
- All possible combinations for z_1 and z_2 from 0 to z_{max1} and 0 to z_{max2} are calculated.

- The binormal cdf is taken, centered at each z value, on the interval $\left[z_{1_i} - \frac{z_{step}}{2}, z_{1_i} + \frac{z_{step}}{2}\right]$ and $\left[z_{2_j} - \frac{z_{step}}{2}, z_{2_j} + \frac{z_{step}}{2}\right]$ as shown in equation 8 below.

$$P\left(z_1, z_2 \middle| x_i, y_j\right) = \int_{z_1 - \frac{z_{step}}{2}}^{z_1 - \frac{z_{step}}{2}} \int_{z_2 - \frac{z_{step}}{2}}^{z_2 + \frac{z_{step}}{2}} \frac{1}{\sqrt{\left(2\pi\sigma_{x,y,z_1}\sigma_{x,y,z_2}\sqrt{1-\rho^2}\right)}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{\left(z_1 - \mu_{x,y,z_1}\right)^2}{2\sigma_{x,y,z_1}{}^2} + \frac{\left(z_2 - \mu_{x,y,z_2}\right)^2}{2\sigma_{x,y,z_2}{}^2} - \frac{2\rho\left(z_1 - \mu_{x,y,z_1}\right)\left(z_2 - \mu_{x,y,z_2}\right)}{\sigma_{x,y,z_1}\sigma_{x,y,z_2}}\right]\right) dz_2 dz_1$$ **Equation S11**

- This value is then used in Equation S7-9 to calculate the appropriate probabilities and information.

7.1.7 Alternative Methods of Information Calculation

In addition to calculating information from the summation of probabilities, one may also calculate information from the entropies. Three equations for entropy are defined below in Equation S12a-c.

$$H(X, Y; Z) = \sum_{z \in Z} \sum_{i=1}^{n} \sum_{j=1}^{m} p\left(x_i, y_j, z\right) \log\left(p\left(x_i, y_j, z\right)\right)$$ **Equation S12a**

$$H(X; Z) = \sum_{z \in Z} \sum_{i=1}^{n} p(x_i, z) \log\left(p(x_i, z)\right)$$ **Equation S12b**

$$H(Z) = \sum_{z \in Z} p_Z(z) \log\left(p_Z(z)\right)$$ **Equation S12c**

Using these entropies, one can arrive at the value for mutual information as shown in Equation S13a-b [8].

$$I(X,Y;Z) = H(X,Y) + H(Z) - H(X,Y,Z)$$ **Equation S13a**

$$I(X;Z) = H(X) + H(Z) - H(X,Z)$$ **Equation S13b**

Aside from ignoring a variable in the mutual information given by I(X,Z), one may eliminate a variable by taking the conditional mutual information given by I_Y(X,Z) in Equation S14.

$$I_Y(X;Z) = I(X,Y;Z) - I(X;Z)$$ **Equation S14**

The conditional mutual information allows us to calculate another value of interest, the interaction information. Interaction information, I(X;Y;Z) Equation S15, describes the amount of information (either redundancy or synergy) given by a set of variables, beyond that which is present in any subset of those variables [3]. A negative value corresponds with a redundancy whereas a positive value corresponds with a synergy.

$$I(X;Y;Z) = I_Y(X;Z) - I(X;Z)$$ **Equation S15**

7.1.8 Information Calculation Verification

Equation S1 and Equation S5 use summation of information values to determine total and mutual information. McGill provides an alternative method in the summation of entropies [3]. These equations, given as Equation S13a, result in nearly identical values of information. The detailed calculations of this are shown in Equation S16-S18.

$$I(X,Y;Z) = H(X,Y) + H(Z) - H(X,Y,Z) = 9.4 + 11.70 - 17.46$$
$$= 3.640\ vs.\ 3.634$$
Equation S16

$$I(X;Z) = H(X) + H(Z) - H(X,Z) = 4.7 + 11.70 - 15.72 = 0.68\ vs.\ 0.68$$
Equation S17

$$I(Y;Z) = H(Y) + H(Z) - H(Y,Z) = 4.7 + 11.70 - 15.72 = 0.68\ vs.\ 0.68$$
Equation S18

7.1.9 Full Effect of Parameters

The effects of all parameter changes for Smad model A are tabulated in Table S3.5-6. Percent change in information was calculated using Equation S19. Absolute change in efficiency was calculated using Equation S20.

$$\frac{I_{new} - I_{standard}}{I_{standard}}$$
Equation S19

$$\frac{I_{new}}{H} - \frac{I_{standard}}{H}$$
Equation S20

Note that for the large majority of parameters, there was little change in information when the parameter was varied symmetrically. Exceptions are discussed in the main article.

Table S3.6. Effect of Asymmetric Parameter Change

		Percent Change in Information (% of Bits)			Absolute Change in Efficiency (%)		
	Fold Change	I(X,Y;Z)	I(X;Z)	I(Y;Z)	$\frac{I(X,Y;Z)}{H(X,Y;Z)}$	$\frac{I(X;Z)}{H(X;Z)}$	$\frac{I(Y;Z)}{H(Y;Z)}$
δ_R	x10	-14.97%	-62.97%	107.18%	-4.96%	-9.06%	15.92%
	x0.1	11.11%	31.72%	-12.25%	5.39%	4.84%	-1.62%
δ_X	x10	-5.22%	-24.09%	34.04%	-1.09%	-3.36%	5.18%
	x0.1	2.55%	9.95%	-2.77%	1.99%	1.64%	-0.23%
γ_A	x10	4.81%	306.52%	-55.28%	2.89%	45.18%	-7.94%
	x0.1	1.66%	-64.08%	307.59%	1.64%	-9.23%	45.34%
γ_B	x10	-0.11%	0.19%	0.24%	0.94%	0.21%	0.22%
	x0.1	-0.14%	-0.04%	0.26%	0.92%	0.17%	0.22%
γ_C	x10	-0.09%	1.36%	-1.03%	0.94%	0.38%	0.03%
	x0.1	0.44%	-8.12%	14.01%	1.16%	-1.01%	2.24%
K_x	x10	15.88%	48.65%	-13.64%	7.28%	7.32%	-1.82%
	x0.1	-14.43%	-65.60%	117.47%	-4.74%	-9.45%	17.43%

When parameters are varied asymmetrically, there was often a loss of information for one signal with a gain in the other. For many parameters, however, this was not a significant change in information.

To demonstrate the effect that changing various parameters has on the dynamic range of the output refer to Figures S3.3-4.

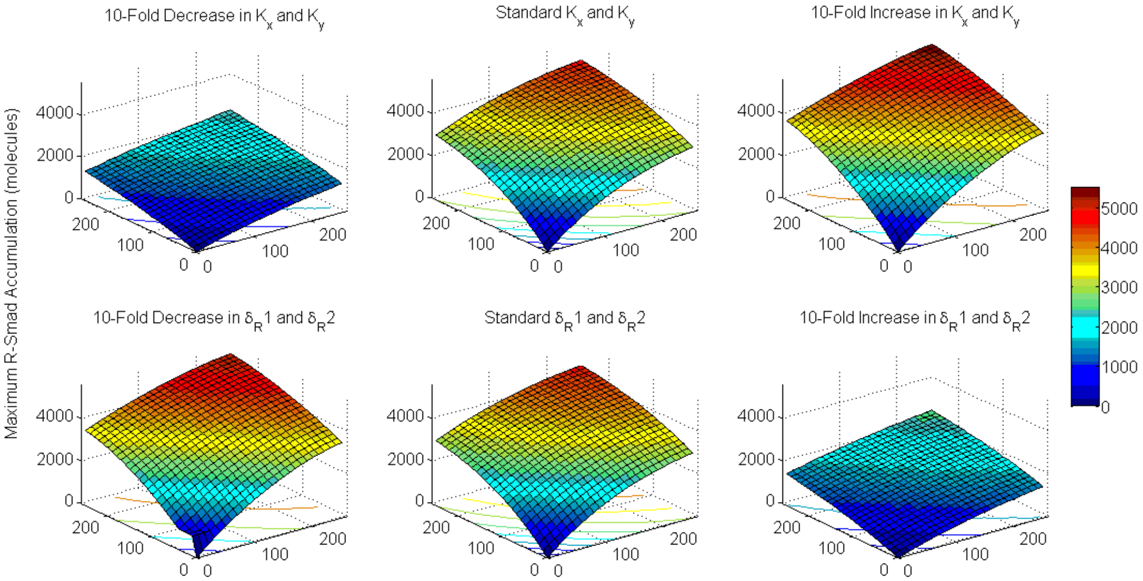


Figure S3.3. **Dynamic Range of Maximal R-Smad Accumulation for Bilateral Fold Changes of K_x {K_y and δ_{R1} {δ_{R2}.**

This figure demonstrates the effect of symmetrically varying the rates of receptor degradation and ligand binding. Note that the panels increase/decrease symmetrically. The x- and y-axes correspond to the initial ligand amount of ligand X and Y respectively. The z-axis is the average maximum accumulation of the output, phosphorylated RSmad.

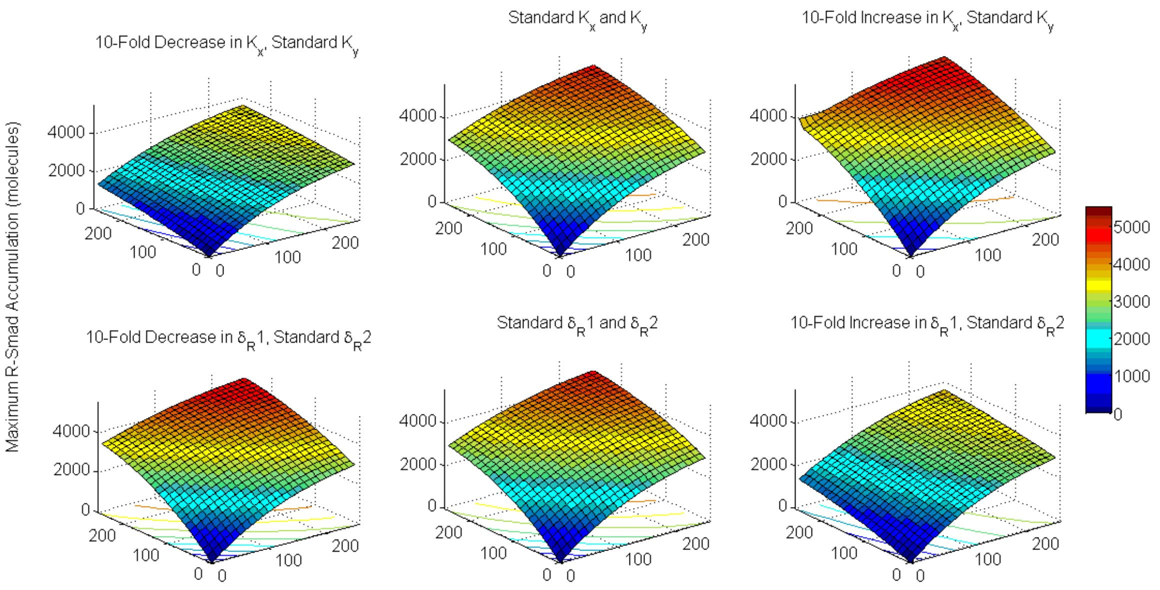


Figure S3.4. **Dynamic Range of Maximal R-Smad Accumulation for Unilateral Fold Changes of K_x & K_y and δ_{R1} & δ_{R2}.**
This figure demonstrates the effect of asymmetrically varying the rates of receptor degradation and ligand binding. Note that the panels become skewed as the rates are increased/decreased asymmetrically for X and Y. The x- and y-axes correspond to the initial ligand amount of ligand X and Y respectively. The z-axis is the average maximum accumulation of the output, phosphorylated RSmad.

7.2 Full Results from Smad Model B

To compare the information between a single output and the bivariate output, information was calculated using only RSmad:p, only RSmad:Co-Smad, and using both, as shown in Table S3.7.

Table S3.7. Model 1 Information and Entropy

	I(X,Y;Z)	I(X;Z)	I(Y;Z)	I_Y(X;Z)	I_X(Y;Z)	I(X;Y;Z)	I(Y;X;Z)
Z= RSmad:p	3.47	0.675	0.675	2.80	2.80	2.13	2.13
Z= RSmad:Co-Smad	3.49	0.675	0.675	2.82	2.81	2.14	2.14
Z=(RSmad:p, RSmad:Co-Smad)	3.52	0.680	0.680	2.83	2.83	2.15	2.15

	H(Z)	H(X)	H(Y)	H(X;Y)	H(X;Z)	H(Y;Z)	H(X,Y;Z)
Z= RSmad:p	9.23	4.70	4.70	9.40	13.25	13.26	15.16
Z= RSmad:Co-Smad	11.26	4.70	4.70	9.40	15.29	15.29	17.17
Z=(RSmad:p, RSmad:Co-Smad)	8.58	4.70	4.70	9.40	12.60	12.60	14.46

7.3 Parameter Sensitivity for Two-Component Model

Two key parameters were tested for their effect on information transfer over a range of cross talk values. The ligand on rate (k_x,k_y) and the ligand off rate (δ_x,δ_y) were varied over several orders of magnitude by increasing and decreasing the rates 10-fold, 1000-fold. As shown in Figure S3.5, the trend of a lack of robustness against cross talk is maintained throughout this range of parameter values.

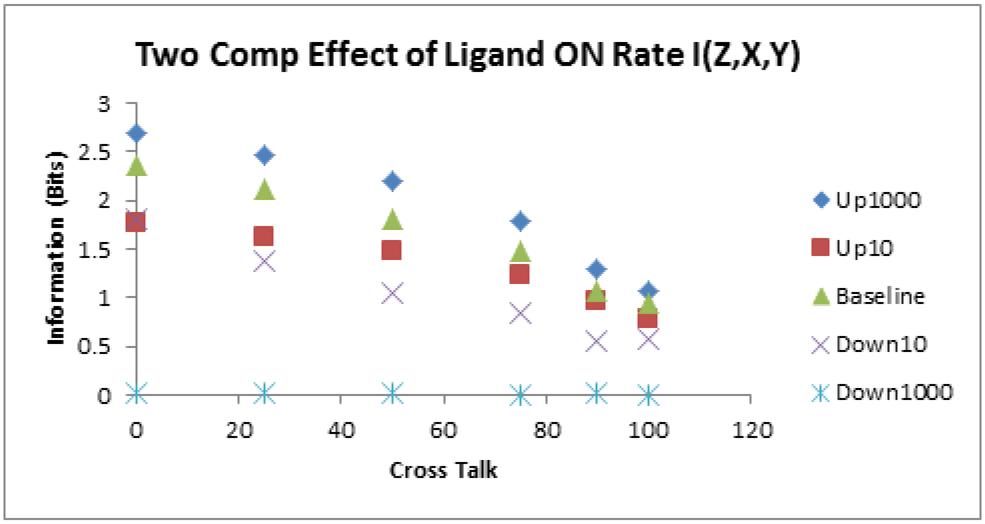


Figure S3.5. Effect of the Ligand On rate, K_x, K_y on information transfer.
The ligand on rate was increased or decreased 10-fold and 1000-fold and tested across a range of cross-talk values. The results from the parameter sensitivity for the ligand off rate (δ_x,δ_y) can be found in Figure S3.6. Similarly, we find that the lack of robustness against cross-talk appears to be insensitive to the parameter.

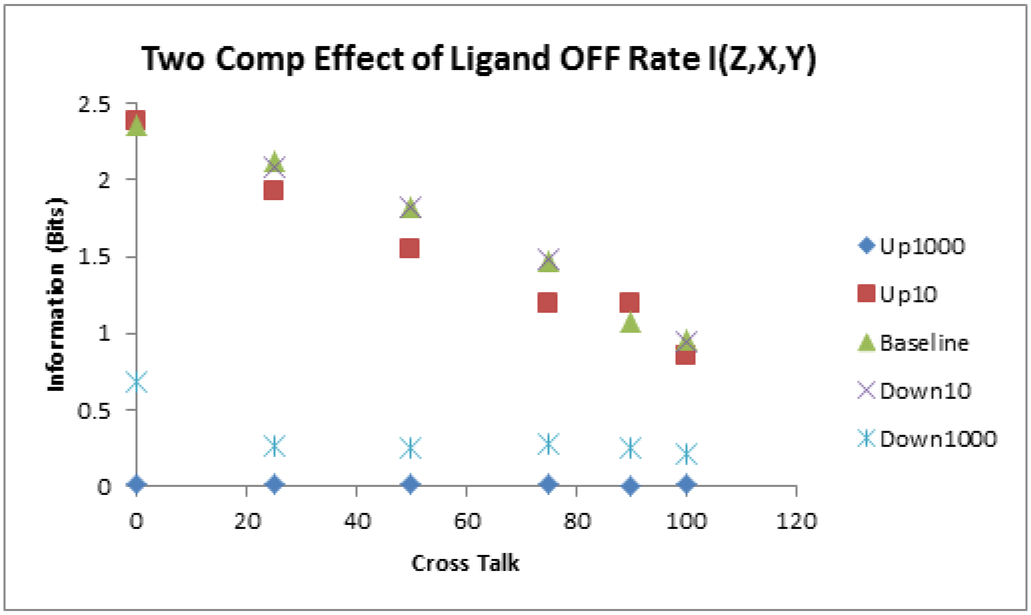


Figure S3.6. Effect of the Ligand Off rate, δ_x, δ_y on information transfer. The ligand off rate was increased or decreased 10-fold and 1000-fold and tested across a range of cross-talk values

7.4 Parameter Sensitivity for Smad Model C

The two key parameters identified above (Full Effect of Parameters) were tested for their effect on information transfer over a range of cross talk values. The ligand on rate (k_x,k_y) and the receptor internalization rate (δ_R) were varied over several orders of magnitude by increasing and decreasing the rates 100-fold or 10-fold and 1000-fold. As shown in **Error! Reference source not found.**, the trend of robustness against cross talk is maintained throughout this range of parameter values.

The results from the parameter sensitivity for the receptor internalization rate (δ_x,δ_y) can be found in **Error! Reference source not found.**. Similarly, we find that the robustness against cross-talk appears to be insensitive to the parameter.

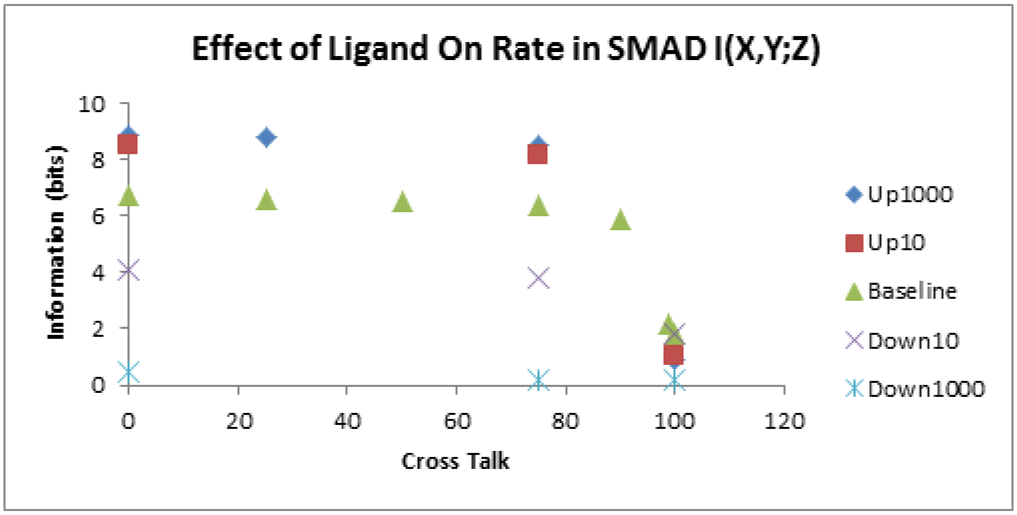


Figure S3.7. Effect of the Ligand On rate, K_x,K_y on information transfer. The ligand on rate was increased or decreased 10-fold and 1000-fold and tested across a range of cross-talk values.

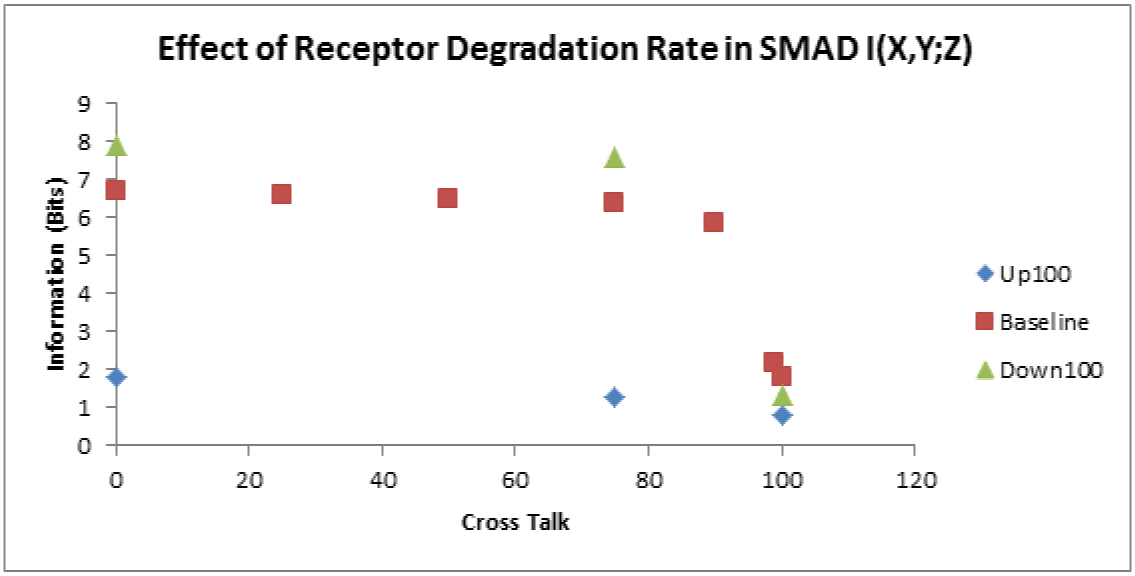


Figure S3.8. Effect of the Ligand Off rate, δ_x, δ_y on information transfer. The ligand off rate was increased or decreased 10-fold and 1000-fold and tested across a range of cross-talk values.

7.5 Adjusted SMAD and Two-Component Models

In order to more accurately assess the effect of dynamic range on the information transfer ability, we adjusted the parameters within each model to approximately match the dynamic range of the opposite model.

7.6 Increased Dynamic Range Two-Component Model

In order to increase the dynamic range of the two-component model to match that of the SMAD model, several parameters were adjusted. These are highlighted in Table S8. Parameters with new values are in bold. A plot of the original SMAD dynamic range and the increased two-component models is shown in Figure S3.9.

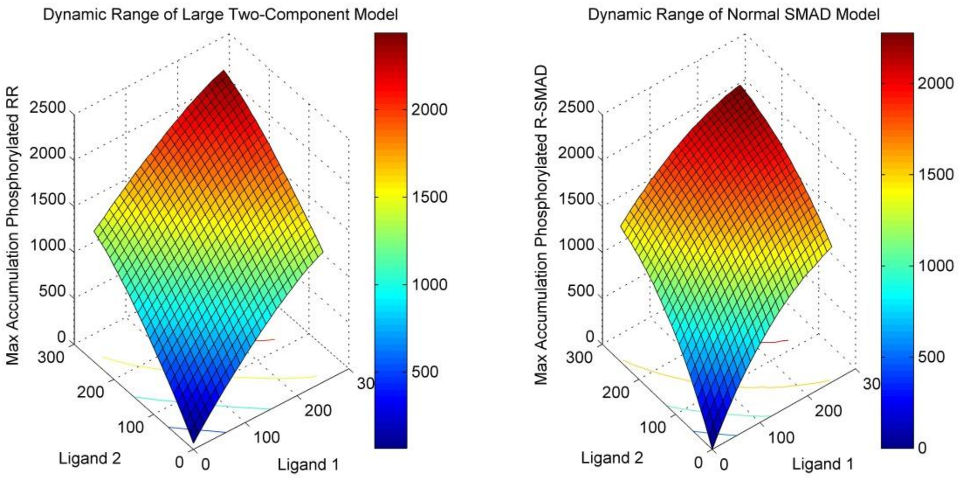


Figure S3.9. The Dynamic Range of the Large Two-Component Model. Falls within the dynamic range of the small smad model.

7.7 Decreased Order of Magnitude SMAD Model

In order to decrease the order of magnitude of the SMAD model to match that of the two-component model, the dephosphorylation rate was increased to 50E-3 from 6.6E-3. A plot of the decreased SMAD order of magnitude and the original two-component models is shown in Figure S3.10.

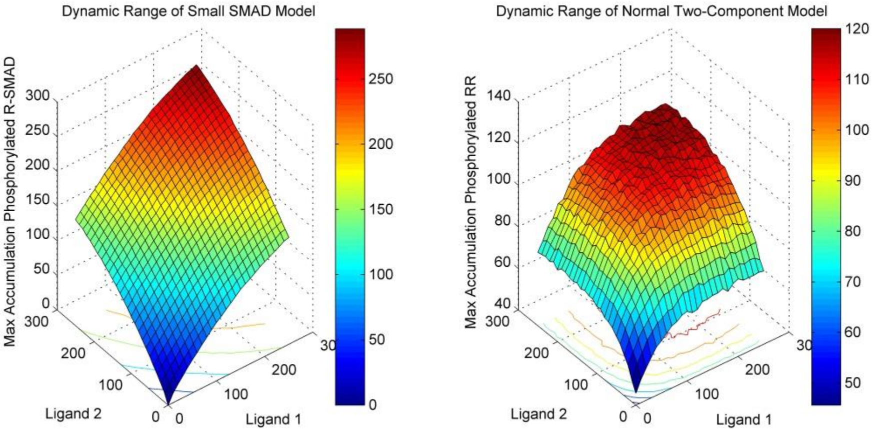


Figure S3.10. The Dynamic Range of the Small SMAD Model. Is of the same order of magnitude as the two-component model.

REFERENCES

[1] Igoshin, O. A., Alves, R., & Savageau, M. A. (2008). Hysteretic and graded responses in bacterial two-component signal transduction. Molecular Microbiology, 68[5], 1196-1215.

[2] Lis, M., Devadas, S., & Chakraborty, A. (2009). Efficient stochastic simulation of reaction-diffusion processes via direct compilation. Bioinformatics, 25(17), 2289-2291.

[3] McGill, W. J. (1954). Multivariate information transmission. Psychometrika, 19(2), 97-116.

[4] Mehta, P., Goyal, S., Long, T., Bassler, B. L., & Wingreen, N. S. (2009). Information processing and signal integration in bacterial quorum sensing. Molecular Systems Biology, 5(325).

[5] Nakabayashi, J., & Sasaki, A. (2009). A mathematical model of the stoichiometric control of Smad complex formation in TGF-beta signal transduction pathway. Journal of Theoretical Biology, 259(2), 389-403.

[6] Reza, F. (1994). An introduction to information theory. New York: Dover.

[7] Schmierer, B., Tournier, A. L., Bates, P. A., & Hill, C. S. (2008). Mathematical modeling identifies Smad nucleocytoplasmic shuttling as a dynamic signal-interpreting system. Proceedings of the National Academy of Sciences, 105(18), 6608-6613.

[8] Srinivasa, S. (2005). A Review on multivariate mutual information. Information Theory Tutorials, University of Notre Dame.

APPENDIX II: SUPPLEMENTARY INFORMATION FOR CHAPTER 4

8.1 Genetic Toggle Switch

8.1.1 Derivation and Dedimensionalization of Genetic Toggle Switch

Let [R$_1$], [R$_2$] be the concentration repressors 1 and 2. Let [L$_{1F}$] and [L$_{2F}$] be the concentration of unbound load 1 and 2 with total load concentration [L$_{1T}$] and [L$_{2T}$], and which can bind R1 and R2 reversibly. Let [C$_1$] and [C$_2$] be the concentration of a load-receptor complex.

$$\left[R_1\right]+\left[L_{1F}\right]\underset{k_{off1}}{\overset{k_{on1}}{\rightleftarrows}}\left[C_1\right]$$

$$\left[R_2\right]+\left[L_{2F}\right]\underset{k_{off2}}{\overset{k_{on2}}{\rightleftarrows}}\left[C_2\right]$$

R$_1$ is produced at a maximal rate of β$_1$ and is repressed by R$_2$. R$_2$ is produced at a maximal rate of β$_2$ and repressed by R$_1$. R$_1$ and R$_2$ degrade as a first order process at rate δ. Thus the differential equations governing [R$_1$] and [R$_2$] are:

$$\frac{d\left[R_1\right]}{dt}=\frac{b_1}{1+\left(\left[R_2\right]/k_2\right)^n}-d\left[R_1\right]-k_{on1}\left[R_1\right]\left[L_{1F}\right]+k_{off1}\left[C_1\right]$$

$$\frac{d\left[R_2\right]}{dt}=\frac{b_2}{1+\left(\left[R_1\right]/k_1\right)^n}-d\left[R_2\right]-k_{on2}\left[R_2\right]\left[L_{2F}\right]+k_{off2}\left[C_2\right]$$

We now de-dimensionalize these equations and define the following parameters. L$_{10}$, L$_{20}$ are total amount of load 1 and 2 respectively.

$$u=\left[R_1\right]/k_1 \qquad v=\left[R_2\right]/k_2 \qquad t=td$$

$$\ell_1 = [L_{1F}]/[L_{1T}] \qquad [C_1] = [L_{1T}] - [L_{1F}] \quad C_1' = [L_{1T}](1 - \ell_1)$$

$$\ell_2 = [L_{2F}]/[L_{2T}] \qquad [C_2] = [L_{2T}] - [L_{2F}] \quad C_2' = [L_{2T}](1 - \ell_2)$$

$$\frac{d dk_1 u}{dt} = \frac{b_1}{1+v^n} - dk_1 u - k_{on1} k_1 u \ell_1 L_{1T} + k_{off1} L_{1T}(1 - \ell_1)$$

$$\frac{d dk_2 v}{dt} = \frac{b_2}{1+u^n} - dk_2 v - k_{on2} k_2 v \ell_2 L_{2T} + k_{off2} L_{2T}(1 - \ell_2)$$

These equations reduce to:

$$\frac{du}{dt} = \frac{b_1/dk_1}{1+v^n} - u - \frac{k_{on1}[L_{1T}]}{d} u \ell_1 + \frac{k_{off1}[L_{1T}]}{dk_1}(1 - \ell_1)$$

$$\frac{dv}{dt} = \frac{b_2/dk_2}{1+u^n} - v - \frac{k_{on2}[L_{2T}]}{d} v \ell_2 + \frac{k_{off2}[L_{2T}]}{dk_2}(1 - \ell_2)$$

We now redefine several parameters and add a basal production rate of u and v, α_1 and α_2:

$$b_1\text{¢}= b_1/dk_1 \qquad k_{on1}\text{¢}= \frac{k_{on1}}{d} \qquad k_{off1}\text{¢}= \frac{k_{off1}}{dk_1}$$

$$b_2\text{¢}= b_2/dk_2 \qquad k_{on2}\text{¢}= \frac{k_{on2}}{d} \qquad k_{off2}\text{¢}= \frac{k_{off2}}{dk_2}$$

The final differential equations are therefore:

$$\frac{du}{dt} = a_1 + \frac{b_1'}{1+v^n} - u - k_{on1}'[L_{1T}]u\ell_1 + k_{off1}'[L_{1T}](1 - \ell_1) \qquad [S1]$$

$$\frac{dv}{dt} = a_2 + \frac{b_2'}{1+u^n} - v - k_{on2}'\left[L_{2T}\right]v\ell_2 + k_{off2}'\left[L_{2T}\right]\left(1 - \ell_2\right) \qquad [S2]$$

$$\frac{d\ell_1}{dt} = - k_{on1}\text{¢}k_1 u\ell_1 + k_{off1}\text{¢}k_1\left(1 - \ell_1\right) \qquad [S3]$$

$$\frac{d\ell_2}{dt} = - k_{on2}\text{¢}k_2 v\ell_2 + k_{off2}\text{¢}k_2\left(1 - \ell_2\right) \qquad [S4]$$

$\alpha_1 = \alpha_2 = 0.2$; $\beta_1' = \beta_2' = 4$; n=3; $k_{on1}' = k_{on2}' = 0.5$; $k_{off1}' = k_{off2}' = 0.5$; $k_1 = k_2 = 1$; $[L_{1T}]$ and $[L_{2T}]$ are variable.

Note that the bifurcation analysis presented later in the Supplementary figures demonstrates that the various models that we study possess at least one stable state, i.e. the Jacobian of the system has eigenvalues with all negative real parts at the fixed points, and hence are asymptotically stable (see for example Theorem 4.6, p121 in “Differential Dynamic Systems” by Meiss). Thus all models discussed can be assumed to show convergence to equilibrium.

8.1.2 The simulation box and concentrations for the stochastic simulations.

We used a volume of 1 μm^3 for the simulation box and the base parameters for simulation of the genetic toggle switch as in the deterministic simulations. This corresponded to small numbers of about 10-20 molecules of the two repressors in the simulation box. The small numbers of molecules led to frequent stochastic events and many spontaneous transitions between the two states. The rate expressions used for the stochastic simulations of the genetic toggle switch are shown in Table S4.5.

8.1.3 Sensitivity to molecule number of the genetic toggle switch

In order to test whether our procedure for constructing the quasi-potential landscape for the genetic toggle switch was robust to larger molecule numbers, we ran the simulation with the average number of molecules in each state about 5 times larger than that reported in the text. As shown in Figure S4, the trends are similar to those reported in the main paper. Note that the number of transitions seen in any block of time is much fewer and hence it takes a significantly longer computational time to actually collect enough data for smooth and accurate plots. However there is no qualitative change in the results due to a larger molecule number.

8.1.4 Parameter Sensitivity in Deterministic Simulations and Alternative Induction Method

8.1.4.1 Alternative Induction Method

We considered a second method of induction that utilizes an inducer that directly reduces the level of a repressor. We derived similar a system of four differential equations listed below:

$$\frac{du}{dt} = a_1 + \frac{b_1'}{1+v^n} - \left(u + \frac{g_3 I_1}{1+I_1}\right) - k_{on1}'\left[L_{1T}\right]u\ell_1 + k_{off1}'\left[L_{1T}\right]\left(1 - \ell_1\right) \qquad \text{[S5]}$$

$$\frac{dv}{dt} = a_2 + \frac{b_2'}{1+u^n} - \left(v + \frac{g_4 I_2}{1+I_2}\right) - k_{on2}'\left[L_{2T}\right]v\ell_2 + k_{off2}'\left[L_{2T}\right]\left(1 - \ell_2\right) \qquad \text{[S6]}$$

$$\frac{d\ell_1}{dt} = - k_{on1}¢k_1 u\ell_1 + k_{off1}¢k_1\left(1 - \ell_1\right) \qquad \text{[S7]}$$

$$\frac{d\ell_2}{dt} = - k_{on2}¢k_2 v\ell_2 + k_{off2}¢k_2\left(1 - \ell_2\right) \qquad \text{[S8]}$$

γ_3 and γ_4 represent the activities of a factor (I_1 or I_2) that degrades R1 or R2, similar to the degradation of λ CI by RecA, modeled as in [1].

Because the qualitative results for both transition time and inducer required to transition did not vary with induction method, we report the results for the second method below in this Supplementary Information.

8.1.4.2 Transition Time

The relationship between amount of load and transition time was found to be linear across a range of parameters and both induction methods. This was true for a range of load binding on rates from K_{on}= 5 to 0.005. Of note is the identification of an optimal K_d for load binding which results in maximal effect on transition time. This can be seen in Figure S4.1. The effects of a low, medium and high K_d is demonstrated in Figure S4.2. We additionally varied β across two orders of magnitude; the relationship between transition time and load was found to be linear as shown in Table S4.1.

8.1.4.3 Inducer Required to Transition

The inducer decay rate, d_{I1}, affects the exponential parameter. As decay rate increases, more inducer is required to transition because the inducer persists in the system for less time. The exponential relationship can actually be written as a function of the decay rate: $Inducer=C*exp(k*d_I*Load)$. This fact shows that the exponential relationship between the amount of inducer required to transition states and the load applied to the system is dependent upon the decay rate of the inducer. This procedure was repeated for both induction methods resulting in similar qualitative results. The results are shown in Tables S4.2 and S4.3.

8.1.5 Effects of a Dynamic Load

We explored the possibility that a load was not present in a constant amount but rather varied as the load was created and degraded. The equations used for a dynamic load are:

$$\frac{du}{dt} = \alpha_1 + \frac{\beta_1'}{1+v^n} - u - k_{on1}'\frac{k_{b1}}{k_{d1}}u\ell_1 + k_{off1}'\frac{k_{b1}}{k_{d1}}\left(c_1\right) \qquad [S9]$$

$$\frac{dv}{dt} = \alpha_2 + \frac{\beta_2'}{1+u^n} - v - k_{on2}'\frac{k_{b2}}{k_{d2}}v\ell_2 + k_{off2}'\frac{k_{b2}}{k_{d2}}\left(c_2\right) \qquad [S10]$$

$$\frac{d\ell_1}{dt} = -k_{on1}'k_1\frac{k_{b2}}{k_{d2}}u\ell_1 + k_{off1}'k_1\frac{k_{b1}}{k_{d1}}\left(c_1\right) + \frac{k_{d1}}{\delta} - \frac{k_{d1}}{\delta}\ell_1 \qquad [S11]$$

$$\frac{d\ell_2}{dt} = -k_{on2}'k_2\frac{k_{b2}}{k_{d2}}v\ell_2 + k_{off2}'k_2\frac{k_{b2}}{k_{d2}}\left(c_2\right) + \frac{k_{d2}}{\delta} - \frac{k_{d2}}{\delta}\ell_2 \qquad [S12]$$

$$\frac{dc_1}{dt} = k_{on1}'k_1u\ell_1 - k_{off1}'k_1c_1 \qquad [S13]$$

$$\frac{dc_2}{dt} = k_{on2}'k_2u\ell_2 - k_{off2}'k_2c_2 \qquad [S14]$$

Where k_{b1} is the creation rate of load 1 and k_{d1} is the degradation rate. K_{eq1}, defined as k_{b1}/k_{d1} was chosen as the de-dimensionalization constant (similar to L_{1T}, L_{2T} used above). As a result, the transition times were plotted against K_{eq} rather than L_T. The default parameter values used were: $k_{d1}=k_{d2}=0.5$; $k_1=k_2=1$; $\delta=1$; k_{b1} and k_{b2} were varied from 0.5 to 50 to cover a range of loading conditions. In addition to the default parameters, we tested the effects of k_d, k_1, k_{on} and

k_{off} on the relationship between load and transition time. Because the transitions between states are not induced until the system has reached a steady state, there was no qualitative effect on the deterministic results. The relationship between K_{eq} for the load and transition time was found to be linear in all parameter regimes. This is shown for the default parameter conditions in Figure S4.3 and for the other parameter conditions in Table S4.4.

8.1.6 Positive Feedback on the Toggle Switch

8.1.6.1 Derivation of Composite Promoter Term

Let *P* be a constitutively active promoter which produces repressor R_1 at rate β:

$$P \xrightarrow{b} P + R_1$$

Let R_1 have positive feedback on *P*:

$$P + R_1 \underset{k_2}{\overset{k_1}{\rightleftarrows}} PR_1 \xrightarrow{r} PR_1 + R_1$$

$$P + R_1 \xleftarrow{k_2} PR_1$$

Let R_1 be a repressor of *P* which binds in n copies:

$$P + nR_2 \underset{k_4}{\overset{k_3}{\rightleftarrows}} PnR_2$$

We assume quasi-steady state:

$$\frac{d\left[PR_1\right]}{dt} = k_1\left[P\right]\left[R_1\right] - k_2\left[PR_1\right] = 0$$

$$\frac{d\left[PR_2\right]}{dt} = k_3\left[P\right]\left[R_2\right]^n - k_4\left[PR_2\right] = 0$$

Therefore:

$$\left[PR_1\right] = \frac{k_1}{k_2}\left[P\right]\left[R_1\right];\ \left[PR_2\right] = \frac{k_3}{k_4}\left[P\right]\left[R_2\right]^n$$

We assume a constant amount of P. From the law of conservation:

$$P_0 = \left[P\right] + \left[PR_1\right] + \left[PnR_2\right]$$

Let $k_1/k_2 = k'$ and $k_3/k_4 = k''$

$$P_0 = \left[P\right] + k'\left[P\right]\left[R_1\right] + k''\left[P\right]\left[R_2\right]^n$$

$$\left[P\right] = \frac{P_0}{1 + k'\left[R_1\right] + k''\left[R_2\right]^n}$$

We now solve for the rate of R1:

$$\frac{d\left[R_1\right]}{dt} = r\left[PR_1\right] - k_1\left[P\right]\left[R_1\right] + k_2\left[PR_1\right] + bP = rk'\left[P\right]\left[R_1\right] + bP$$

$$\frac{d\left[R_1\right]}{dt} = \left[P\right]\left(rk'\left[R_1\right] + b\right) = \frac{P_0 rk'\left[R_1\right] + P_0 b}{1 + k'\left[R_1\right] + k''\left[R_2\right]^n}$$

Let ρ'=P$_0$ρk' and β' = P$_0$β

$$\frac{d\left[R_1\right]}{dt} = \frac{r'\left[R_1\right] + b'}{1 + k'\left[R_1\right] + k''\left[R_2\right]^n}$$

This yields the positive feedback term in Eq. 10 in the text.

8.1.6.2 Strength of Positive Feedback

To assess the effects of a positive feedback on repressor 1, we tested various values of parameter *ρ*, the strength of positive feedback. First note that the positive feedback moiety, even in the presence of the load, does not abrogate bistability, unless *ρ* is very large, as shown in Figure S4.5. We then tested the probability distribution functions for the toggle with positive feedback without and with a load. The results of this are shown in Figure S4.6. When the positive feedback is 0, the probability distribution function for R1 and R2 is perfectly balanced. As the strength of positive feedback increases from 0 to 5, the pdf is increasingly skewed to R1.

When ρ=5, the effects of the positive feedback are so strong that the system never switches stochastically into R2. As shown in the paper, this effect may be overcome by increasing the load on R1.

8.1.6.3 Effect of Positive Feedback on Transition Time

Even in the case of a positive feedback, the relationship between transition time and load remains linear. We explored the effect of transition time when the positive feedback was applied to the R1 and R2. In all cases, the relationship was linear. This is shown in Supplementary Figure S4.5.

8.2 Motivations of the Ras System Model

8.2.1 Assumptions underlying the Ras Model:

The model we used for the Ras-Kinase system is mainly adopted from the minimal model of the Ras Switch proposed by Das et. al. [1] In the following section, we briefly discuss the underlying assumptions of the minimal model of Ras Switch:

1. SOS: As in [1], only SOS (Son of Sevenless) family of Ras Guanine Nucleotide Exchange Factors (GEFs) is included in the model. The RasGRP (Ras Guanine Nucleotide Release Protein) family (including RasGRP1 and RasGRP2) which are also GEFs are not included. SOS is ubiquitously distributed, while RasGRP family is restricted to the nervous and hematopoietic systems.
2. SOScat: Not all the domains of SOS are taken into consideration in this model. Cdc25 and REM, together named as SOScat, are only two included, which are essential domains for GEF catalytic activities. The domains flanking SOScat, both N-terminal and C-terminal, are shown to be inhibitory to GEF activities. *In vivo,* when SOS is recruited to

the plasma membrane, the resulting conformational changes release this inhibition. In this minimal model this inhibiting effect is not modeled. For our purpose of investigating the effects of adding loads to the positive feedback based bistable Ras switch, we also only consider SOScat in our model.

3. SOS basal GEF activity: The original GEF activity of SOS is very low. However, its GEF activity is strongly influenced by the allosteric pocket in REM domain. When RasGDP binds to this pocket a 5-fold increase is observed in its GEF activity, while binding of RasGTP to this site results in an increase of 75 times. Based on the main aim of the paper, we also choose to neglect the original GEF activity of SOScat. However, we also tested this assumption by including this basal GEF activity in the Ras model, whose behaviors show no qualitative differences with the model we described in the main text (data not shown).
4. Intrinsic GTPase activity of Ras: intrinsic GTPase activity of Ras is relatively low. Proteins we have generically called RasGAPs are constitutively present that promote Ras deactivation from RasGTP into RasGDP. For simplicity the intrinsic GTPase activity of Ras is neglected and the enhanced deactivation of RasGTP by RasGAP is modeled as an enzymatic reaction.
5. Truncated Raf cascade: after RasGTP binds to Raf *in vivo*, Raf will be phosphorylated and activated by the RasGTP:Raf complex. Then the activated Raf proteins activate the RAF-MEK-ERK-CD69 pathway. However, for the purpose of our study here, Raf is simplified into only a binding partner of RasGTP. Thus, the downstream phosphorylation and activation steps are not considered.

6. Since we are interested in the short-term behavior of the system, no synthesis or degradation dynamics is considered in our model, i.e. total amounts of all primary molecules (SOScat, Ras protein, RasGAP, Raf) are conserved. Note that experiments show Ras activation peaking in one or two minutes after activation [1,2].
7. All enzymatic reactions are modeled by sequential reactions of enzyme (E) and substrate (S) firstly bind together to form enzyme:substrate complex (ES) with a reaction rate constant of $k_{on(i)}$, then the complex disassociates reversibly with $k_{off(i)}$ or produces the product (P) irreversibly with $k_{cat(i)}$. These reactions are shown schematically as:

$$E + S \overset{kon(i),koff(i)}{\leftrightarrow} ES \overset{kcat(i)}{\rightarrow} P$$

8.2.2 Reactions modeled in Ras model

Based on the abovementioned assumptions, all reactions considered in our model are listed in Table S6. In particular, [R1] and [R2] describe the allosteric pocket reactions of binding and unbinding reactions between RasGDP/RasGTP and the allosteric pocket in SOS REM domain. [R3] and [R4] describe the reactions catalyzed by GEF pocket of SOScat with allosteric pocket occupied by RasGTP and RasGDP, indicated by SOS(RasGTP) and SOS(RasGDP) correspondingly. [R5] describes the RasGTP deactivation reaction into RasGDP enhanced by RasGAP. The last but not least, [R6] describes the binding and unbinding of RasGTP and Raf. An underlying assumption here is the protection model, where RasGTP is assumed to be free from deactivation of RasGAP after being bound to Raf. We refer the reader to a later section where this assumption is relaxed.

8.3 ODE formulation

To be more general, we use Law of Mass Action (LMA) to model all the rates of reactions listed in Supplementary Table S4.6. Then, the following set of ODEs is achieved for the changing rate of each of the involving species, which we will call the LMA model. For the following sections, we use the following notations for the species involved in the system:

$$x_1 \equiv \left[SOScat\right];\ x_2 \equiv \left[RasGDP\right];\ x_3 \equiv \left[RasGTP\right];\ x_4 \equiv \left[SOScat\left(RasGDP\right)\right];$$

$$x_5 \equiv \left[SOScat\left(RasGTP\right)\right];\ x_6 \equiv \left[SOScat\left(RasGDP\right):RasGDP\right];$$

$$x_7 \equiv \left[SOScat\left(RasGTP\right):RasGDP\right];\ x_8 \equiv \left[RasGAP\right];\ x_9 \equiv \left[RasGAP:RasGTP\right];$$

$$x_{10} \equiv \left[Raf\right];\ x_{11} \equiv \left[RasGTP:Raf\right]$$

$$\frac{dx_1}{dt} = -k_{on1}x_1x_2 + k_{off1}x_4 - k_{on2}x_1x_3 + k_{off2}x_5 \tag*{[S15]}$$

$$\frac{dx_2}{dt} = -k_{on1}x_1x_2 + k_{off1}x_4 - k_{on3}x_2x_5 + k_{off3}x_7 - k_{on4}x_2x_4 + k_{off4}x_6 + k_{cat5}x_9 \tag*{[S16]}$$

$$\frac{dx_3}{dt} = -k_{on2}x_1x_3 + k_{off2}x_5 + k_{cat3}x_7 + k_{cat4}x_6 - k_{on5}x_3x_8 + k_{off5}x_9 - k_{on6}x_3x_{10} + k_{off6}x_{11} \tag*{[S17]}$$

$$\frac{dx_4}{dt} = k_{on1}x_1x_2 - k_{off1}x_4 - k_{on4}x_2x_4 + k_{off4}x_6 + k_{cat4}x_6 \tag*{[S18]}$$

$$\frac{dx_5}{dt} = k_{on2}x_1x_3 - k_{off2}x_5 - k_{on3}x_2x_5 + k_{off3}x_7 + k_{cat3}x_7 \tag*{[S19]}$$

$$\frac{dx_6}{dt} = k_{on4}x_2x_4 - k_{off4}x_6 - k_{cat4}x_6 \tag*{[S20]}$$

$$\frac{dx_7}{dt} = k_{on3}x_2x_5 - k_{off3}x_7 - k_{cat3}x_7 \qquad [S21]$$

$$\frac{dx_8}{dt} = - k_{on5}x_3x_8 + k_{off5}x_9 + k_{cat5}x_9 \qquad [S22]$$

$$\frac{dx_9}{dt} = k_{on5}x_3x_8 - k_{off5}x_9 - k_{cat5}x_9 \qquad [S23]$$

$$\frac{dx_{10}}{dt} = - k_{on6}x_3x_{10} + k_{off6}x_{11} \qquad [S24]$$

$$\frac{dx_{11}}{dt} = k_{on6}x_3x_{10} - k_{off6}x_{11} \qquad [S25]$$

The followings are conservation laws for primary molecules (SOS, Ras, GAP and Raf) involved in the system.

$$SOS_T = x_1 + x_4 + x_5 + x_6 + x_7 \qquad [S26]$$

$$Ras_T = x_2 + x_3 + x_4 + x_5 + 2x_6 + 2x_7 + x_9 + x_{11} \qquad [S27]$$

$$GAP_T = x_8 + x_9 \qquad [S28]$$

$$Raf_T = x_{10} + x_{11} \qquad [S29]$$

We can recover the equations used in Ref. [1] with: 1) classic Pseudo Steady State Assumption (PSSA) for all the time derivatives of enzyme-substrate complexes; 2) defining Michealis constants as $K_{(i)M}$ = ($k_{off(i)}$ + $K_{cat(i)}$)/$k_{on(i)}$; 3) Approximations of conservation law by ignoring enzyme-substrate complexes under PSSA. Then the enzymatic reaction rates can be simplified into classical Michealis Menten Kinetics and the overall set of ODEs simplified as:

$$x_1 \equiv \left[SOScat\right];\ x_2 \equiv \left[RasGDP\right];\ x_3 \equiv \left[RasGTP\right];\ x_4 \equiv \left[SOScat\left(RasGDP\right)\right];$$

$$x_5 \equiv \left[SOScat\left(RasGTP\right)\right];\ x_6 \equiv \left[SOScat\left(RasGDP\right):RasGDP\right];$$

$$x_7 \equiv \left[SOScat\left(RasGTP\right):RasGDP\right];\ x_8 \equiv \left[RasGAP\right];\ x_9 \equiv \left[RasGAP:RasGTP\right];$$

$$x_{10} \equiv \left[Raf\right];\ x_{11} \equiv \left[RasGTP:Raf\right]$$

$$\frac{dx_1}{dt} = -k_{on1}x_1x_2 + k_{off1}x_4 - k_{on2}x_1x_3 + k_{off2}x_5 \qquad \text{[S30]}$$

$$\frac{dx_3}{dt} = -k_{on2}x_1x_3 + k_{off2}x_5 + \frac{k_{cat3}x_2x_5}{K_{m3}+x_2} + \frac{k_{cat4}x_2x_4}{K_{m4}+x_2} - \frac{k_{cat5}GAP_Tx_3}{K_{m5}+x_3} - k_{on6}x_3x_{10} + k_{off6}x_{11} \qquad \text{[S31]}$$

$$\frac{dx_5}{dt} = k_{on2}x_1x_3 - k_{off2}x_5 \qquad \text{[S32]}$$

$$\frac{dx_{11}}{dt} = k_{on6}x_3x_{10} - k_{off6}x_{11} \qquad \text{[S33]}$$

The total molecular numbers of SOS, Ras and Raf are conserved in the PSSA system resulting in three additional conservation equations:

$$SOS_T = x_1 + x_4 + x_5 \qquad \text{[S34]}$$

$$Ras_T = x_2 + x_3 + x_4 + x_5 + x_{11} \qquad \text{[S35]}$$

$$Raf_T = x_{10} + x_{11} \tag{S36}$$

8.3.1 Simulation box and parameters

A quasi-two dimensional simulation box, similar to the one used by Das et al. [1] was utilized. Our simulation box is a 2μm by 2 μm surface with a height of 1.7 nm. Every molecular species is assumed to be well-mixed in this box.

Reaction rate constants were referenced from [1] and [3]. In our analysis, molecular concentrations were converted to molecular numbers in the simulation box for both deterministic and stochastic analyses. Thus, reaction rate constants need to be converted accordingly. For our simulation box, all the reactions are assumed to happen in the membrane area, which can be considered as two-dimensional. Therefore, the reaction rate constants need conversions not only from concentrations to molecular numbers, but also from 3D to 2D. Based on the simulation box used, a factor of 0.941/4 is used for the conversion from 3D rate constants with unit of (μM^{-1} s^{-1}) to 2D rate constants with unit of (Molecules^{-1} s^{-1}). All the parameters we used in our deterministic and stochastic studies are listed in Supplementary Table 7 with both 3D and 2D values.

For all the following studies, 75 molecules of Ras, 6 molecules of RasGAP were used unless otherwise indicated.

8.3.2 Results for both cases of LMA and PSSA

Results for both the case of LMA and PSSA are shown in Figure 9 (main text) and Figure S7 correspondingly. For the case of LMA, the red line shows the bifurcation analysis results for the Ras system without Raf inside. A bistable regime is observed. While adding more Raf molecules into the system, both limit points are shifting to the right, bistable regime is decreasing and maximal excitable level of RasGTP is decreasing. When more than 21 molecules of Raf are added, as shown by the curve of “Total Raf = 25”, the bistable regime totally diminishes and for all values of SOScat only one monostable point of the system remains.

Supplementary Figure S4.7 shows the case of PSSA, similar pattern of the changes in bistable region can be observed as shown in Figure 4.9 (main text). The only difference between these two cases is for the case of PSSA slightly more Raf molecules are needed to achieve same effect.

To directly examine the effects of adding different amount of Raf into the Ras activation system for both LMA and PSSA models, we also carried out bifurcation analyses with total number of Raf, i.e. Raf_T as primary parameter, which are shown in Supplementary Figure S4.8 and S4.9 correspondingly. Both Supplementary Figure S4.8 and S4.9 start with bistable region when there is no Raf in the system as we can predict. With the increase of total number of Raf in the system, the values of “high” steady state decrease together with increase in the values of the unstable steady state. This results in vanishing of both steady states and only one monostable region after Raf_T reaches a certain threshold. Again, similar patterns are observed for both cases of LMA and PSSA, and the only difference between them is the scale.

8.3.3 Parameter Sensitivity Analysis (PSA)

For PSA of the Ras Minimal Model, we refer the reader to Das et. al. [1]. For our main purpose of investigating the load to the Ras Switch, we varied the values of k_{on6} and k_{off6} and check their influences on the bistability of the system. We first maintained the same ratio of k_{on6} to k_{off6}, then we changed this ratio and check individual influences.

When keep the ratio between k_{on6} and k_{off6} the same and vary absolute values of k_{on6} and k_{off6}, no difference in bifurcation diagram is noticed (data not shown). This means there are no changes in steady state values if the ratio between these two parameters is maintained.

Then k_{on6} and $k_{off6}$are varied separately. As shown in Figure S10, increase of k_{off6} results in left shifts of both limit points, increase in bistable regime and increase in maximal RasGTP activation level. Qualitative features of bistability are maintained. Decrease of k_{off6} results in right shift of both limit points, increase in unstable bistable regime and decrease in maximal RasGTP activation level. Qualitative features of bistability are also maintained.

Supplementary Figure S4.11 shows the results of an increase in k_{on6} value. This results in reverse effects as shown in Supplementary Figure S4.10. Noticeably, increase k_{off6} by 10 times has exactly the same effects as decrease k_{on6} by 10 times and *vice versa*. This indicates the key player of k_{on6} and k_{off6} in the Ras system is the value of their ratio and is additional assurance that no changes would be observed when the ratio between these two parameters is maintained.

8.4 Discussion on protection assumption with a toy toggle switch and Ras Model

8.4.1 Assumption of Protection Model

For both the toggle switch and the Ras Switch, we assumed that the output molecule is protected when bound with the Load. For the toggle switch model, we assumed the repressor proteins are free from first order degradation when bound in the repressor-load complex; for the Ras Switch model, we assumed RasGTP is relieved from enhanced GTPase activity by RasGAP in its complex form with Raf. We show the effects of removing this assumption in the main text. Here we report additional data.

The Hill function form of the genetic toggle we have used does not allow for explicit binding of the repressors with the promoters they repress. We also considered the question whether allowing the repressor to decay when bound with the promoter (note, not the load) would have an effect on the system. To elucidate this point, a toy model is proposed to make the lumped processes in the Toggle Switch model more explicit. We present below an LMA based model for the toggle switch which we use to test whether allowing the *decay of the repressor when bound to the promoter it represses* can has any effect on the system.

8.4.2 Toy Toggle Switch model

For this model, we construct a demonstrative classical toggle switch similar to the one discussed in the text, but using LMA. In this model system, two equivalent repressors are expressed by corresponding genes. Inactive repressors monomers then become activated after a trimerization process. Activated repressor trimmers can then bind to corresponding promoters and repress the transcriptions of the other repressor monomer. Without repressor bound to promoters, genes can be transcribed at full rate.

8.4.2.1 Assumptions

1. Several assumptions were made for this model to both meet our purpose and maintain it in an intuitive form.
2. Two sides of the toggle switch are identical, i.e. same reactions involved and same parameters for same reactions.
3. Since we are interested in the steady state behaviors of the model system, transcriptional and translational processes are lumped together into one overall reaction and assumed to happen immediately without delay.
4. Promoter values are approximated as continuous concentrations rather than more appropriate discrete number of sites.
5. "Leaky" transcriptions of the promoters when they are bound by corresponding repressors are ignored.
6. Repressor monomers and trimers are identical in their degrading dynamics, thus same degrading parameters are used.

Reactions included in this toggle switch model were then formulated and listed in Supplementary Table 4.8. Particularly, [P1] and [P8] describe the trimerization reactions of inactive repressor monomers ($R_{(i)}$) into active repressor trimers ($AR_{(i)}$). [P2] and [P9] describe the binding and unbinding reactions of active repressor trimers to corresponding promoters ($Pro_{(i)}$) to form repressor:promoter complex ($AR_{(i)}$: $Pro_{(i)}$). [P3] and [P10] describe the "ON state" of the promoters without being bound by repressor, which directly give birth to repressor monomers. [P4] and [P11] will be used to test the protection model, which describes the degradations of repressor trimers when they are bound to promoters. Degradation rate constants of this reaction will be set to zero for protection assumption as control and set equal to other degradation rate constant for our purpose. [P5], [P6], [P12] and [P13] are degradation reaction of both repressor monomers and trimers. [P7] and [P14] describe binding and unbinding reactions between repressor monomers and load molecules ($L_{(i)}$) to form repressor:load complex ($R_{(i)}$:$L_{(i)}$).

8.4.2.2 ODE model

Following notations are used for the toggle switch model:

$$x_1 \equiv \left[R_1\right];\ x_2 \equiv \left[AR_1\right];\ x_3 \equiv \left[Pro_1\right];\ x_4 \equiv \left[AR_1 : Pro_1\right];\ x_5 \equiv \left[L_1\right];\ x_6 \equiv \left[R_1 : L_1\right]$$

$$x_7 \equiv \left[R_2\right];\ x_8 \equiv \left[AR_2\right];\ x_9 \equiv \left[Pro_2\right];\ x_{10} \equiv \left[AR_2 : Pro_2\right];\ x_{11} \equiv \left[L_2\right];$$

$$x_{12} \equiv \left[R_2 : L_2\right]$$

We use Law of Mass Action to formulate all the reaction rates and achieve the following time dependent ODEs for each species:

$$\frac{dx_1}{dt} = -3k_1x_1^3 + 3k_2x_2 + \alpha_1x_9 - k_7x_1 - k_{on1}x_1x_5 + k_{off1}x_6 \qquad [S37]$$

$$\frac{dx_2}{dt} = k_1x_1^3 - k_2x_2 - k_3x_2x_3 + k_4x_4 - k_6x_2 \qquad [S38]$$

$$\frac{dx_3}{dt} = -k_3x_2x_3 + k_4x_4 + k_5x_4 \qquad [S39]$$

$$\frac{dx_4}{dt} = k_3x_2x_3 - k_4x_4 - k_5x_4 \qquad [S40]$$

$$\frac{dx_5}{dt} = -k_{on1}x_1x_5 + k_{off1}x_6 \qquad [S41]$$

$$\frac{dx_6}{dt} = k_{on1}x_1x_5 - k_{off1}x_6 \qquad [S42]$$

Since the toggle switch is symmetric, ODEs for the other side are identical except for indexes of variables and parameters thus not presented here. Also governing this system is the conservation of molecule numbers of promoters:

$$Pro_{1T} = x_3 + x_4 \qquad [S43]$$

To generate Figure S12, following parameter values are used:

$$k_1 = k_8 = 0.3,\ k_2 = k_9 = 10,\ k_3 = k_{10} = 0.6,\ k_4 = k_{11} = 1,\ k_5 = k_{12} = 0.1,$$

$$k_6 = k_{13} = 0.1,\ k_7 = k_{14} = 0.1,\ k_{on} = 0.5,\ k_{off} = 0.2.$$

8.4.2.3 Results

Supplementary Figure S4.12 shows the differences between the system with protection of repressor molecules from degradation when bound to promoter (note: not the load) and the one without. If the protection is not included, a very minor increase in the bistable region can be observed with right shift of upper limit point and left shift of lower limit point.

8.4.3 Toggle Switch without and with Positive Feedback Motif

8.4.3.1 Modifications to original models

If the protection assumption is released for the toggle switch model, two more reactions should be added into the reaction system.

$$C_1 \xrightarrow{d} L_{1F}$$

$$C_2 \xrightarrow{d} L_{2F}$$

Corresponding changes to de-dimensionalized ODEs should also be made:

$$\frac{dl_1}{dt} = - k_{on1}{}' k_1 u l_1 + k_{off1}{}' k_1 \left(1 - l_1\right) + \left(1 - l_1\right) \qquad [S44]$$

$$\frac{dl_2}{dt} = - k_{on2}{}' k_2 v l_2 + k_{off2}{}' k_2 \left(1 - l_2\right) + \left(1 - l_2\right) \qquad [S45]$$

When assuming steady state for the entire system, all the terms introduced by load molecules cannot be cancelled out as in the case of protection model. Thus, influences to steady-state behaviors by adding load molecules to the Toggle Switch system should be expected.

8.4.3.2 Results

Figure 4.2A shows the steady-state effects of adding increasing number of load molecules to both sides of the original genetic toggle switch. Without load molecule, the system is bistable with two stable steady states and one unstable steady state as predicted. With the increase of number of load molecules, all these three steady state become closer and finally meet together at certain level of L_T. Then two steady states vanish and only one stable steady state left. Figure 4.2B shows the steady-state effects of adding increasing number of load molecules to R1 side (L_{1T}) of the original genetic toggle switch. Without any load molecule in the system, the toggle switch is bistable with two stable steady states and one unstable steady state as predicted. Adding increasing number of load molecules results in becoming closer between the upper stable steady state and the unstable steady state. At certain threshold of L_{1T} these two steady states meet and vanish, with only the lower stable steady state left.

Supplementary Figure S4.13 shows effects of adding load molecules to the toggle switch with positive feedback on one side when the protection assumption is released. Interestingly, adding same amount of load molecules to different sides also cause different responses from the system. Adding load to R1 side results in increase of bistable regime and adding to R2 side results in decrease of bistable regime. When equal amount of loads added into both sides, the bistable regime is increased but to an extent smaller than adding to R1 side along. Increase in R1 level is much faster in this case than in Figure 2 due to the positive feedback loop.

8.4.3.3 Transition times in the absence of protection

The relationship between transition times and load is altered when the protection of a lower or absent decay rate of the repressor from the load complex. As predicted from Figure 2, the system with a one-sided load loses bistability with a load of 3.4; with a both sided load the system loses bistability with a load of 11. Thus we tested the transition times of the system within this regime. As shown in Figure 4, when a load is applied to the same side, there is a positive linear relationship between rise time and load, but a negative linear relationship with decay time. Conversely, a load applied to the opposite side results in a negative linear relationship between rise time and load, but a positive linear relationship with decay time. A load applied to both sides result in a negative linear relationship for both rise time and decay time. This result is discussed in the main text.

8.4.4 Ras Model

For Ras model, protection model is also assumed implicitly, since RasGTP is free from GTPase activities after binding to Raf. In this section, we examine potential influences caused by this assumption. For the Ras Switch, we modified the original set of ODE’s by including the RasGTP:Raf complexes as an equivalent substrate of RasGAP.

8.4.4.1 Modifications to original model

If RasGTP can still be deactivated into RasGDP by RasGAP in the complex form with Raf, one more reaction should be included into the Ras System:

$$\text{RasGAP+RasGTP:Raf} \xleftrightarrow[koff5]{kon5} \text{RasGAP:RasGTP:Raf} \xrightarrow{kcat5} \text{RasGAP+RasGDP+Raf}$$

$$\text{RasGAP} + \text{RasGTP} : \text{Raf} \overset{\text{kon5,koff 5}}{\leftrightarrow} \text{RasGAP} : \text{RasGTP} : \text{Raf} \overset{\text{kcat5}}{\rightarrow} \text{RasGAP} + \text{RasGDP} + \text{Raf}$$

$$\text{RasGAP} + \text{RasGTP} : \text{Raf} \overset{\text{kon5,koff 5}}{\leftrightarrow} \text{RasGAP} : \text{RasGTP} : \text{Raf} \overset{\text{kcat5}}{\rightarrow} \text{RasGAP} + \text{RasGDP} + \text{Raf}$$

The same reaction rate constants are assumed for the new reaction as free RasGTP de-activation. One more species is introduced into this system, i.e. $x_{12} \equiv \left[RasGAP : RasGTP : Raf \right]$. Based on this new reaction, several modifications should also be made for the ODEs system including adding more terms into x_2, x_8, x_{10} and x_{11} equations and add a new equation of x_{12}.

$$\frac{dx_2}{dt} = - k_{on1} x_1 x_2 + k_{off1} x_4 - k_{on3} x_2 x_5 + k_{off3} x_7 - k_{on4} x_2 x_4 + k_{off4} x_6 + k_{cat5} x_9 + k_{cat5} x_{12} \qquad [S46]$$

$$\frac{dx_8}{dt} = \dagger - k_{on5} x_3 x_8 + k_{off5} x_9 + k_{cat5} x_9 - k_{on5} x_8 x_{11} + k_{off5} x_{12} + k_{cat5} x_{12} \qquad [S47]$$

$$\frac{dx_{10}}{dt} = - k_{on6} x_3 x_{10} + k_{off6} x_{11} + k_{cat5} x_{12} \qquad [S48]$$

$$\frac{dx_{11}}{dt} = k_{on6} x_3 x_{10} - k_{off6} x_{11} - k_{on5} x_8 x_{11} + k_{off5} x_{12} \qquad [S49]$$

$$\frac{dx_{12}}{dt} = k_{on5} x_8 x_{11} - k_{off5} x_{12} - k_{cat5} x_{12} \qquad [S50]$$

Modifications are also needed for conservation laws:

$$Ras_T = x_2 + x_3 + x_4 + x_5 + 2x_6 + 2x_7 + x_9 + x_{11} + x_{12} \qquad [S51]$$

$$GAP_T = x_8 + x_9 + x_{12} \qquad [S52]$$

$$Raf_T = x_{10} + x_{11} + x_{12} \qquad [S53]$$

8.4.4.2 Results

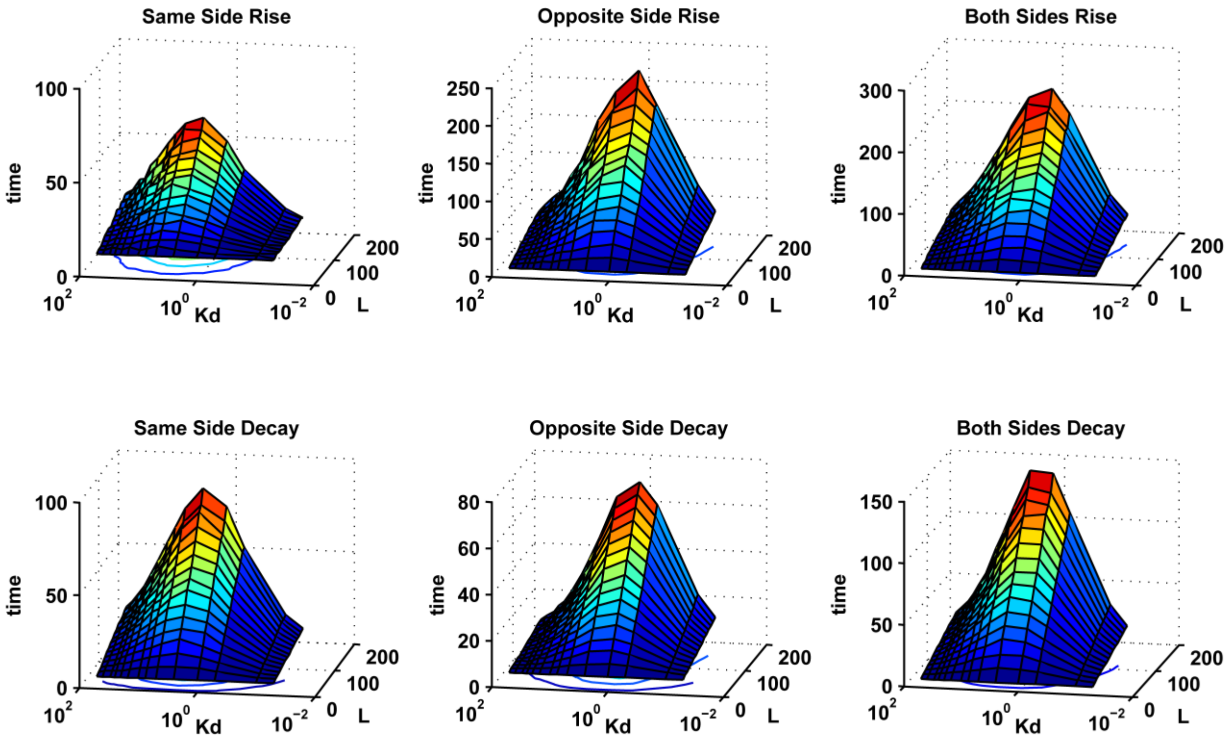
Figure S4.14 shows the changes of bifurcation curve with different numbers of Raf molecules (5, 15, 25 and 200) added into the system in logarithmic scale respectively. A similar pattern of decreases in bistable region and finally elimination of the bistable region as reported in the Figure 9 in main text is still observed but with a more complicated dynamics. Differences between the case without protection model and the one in main text as with protection model will be discussed as follows.

As shown in Figure 4.9 in the main text, where the protection model is included, maximal activation level of RasGTP is always decreasing with increase in Raf molecules added into the system. While in Supplementary Figure S4.14 reported here, maximal activation level of RasGTP first increase (as for the case of “Total Raf=5”) and then decrease (comparing the case of “Total Raf=25” to “Total Raf = 15”) with adding more Raf molecules.

Elimination of bistable region happens with much more Raf molecules. Even though the bistable region already decreased a lot after adding 25 molecules, the left bistable region needs much more Raf molecules to eliminate. Even with 150 Raf molecules, a tiny bistable region still exists. After around 200 molecules of Ras added, the bistability is abrogated.

Decrease in bistable region in Figure 4.9 in the main text is a result of right shifts of both fixed points with a faster rate of shift for the upper fixed points. While the decrease in Supplementary Figure S4.14 is a result of the leftward shift of both fixed points with a faster shift rate for the lower fixed point.

8.5 Figures


Same Side Rise

Opposite Side Rise

Both Sides Rise

Same Side Decay

Opposite Side Decay

Both Sides Decay

Figure S4.1. Surface plots showing response times of the simple genetic toggle switch with changes in load (L) and changes in the dissociation constant (Kd) of binding with load. The units of L and Kd are (molecules/µm3). The z-axis measures the response time indicated in the title. “Same Side Rise” and “Same Side Decay” refers to the rise time and decay time when the load is on the same side as the repressor whose concentration is increasing. “Opposite Side Rise” and “Opposite Side Decay” refers to the rise time and the decay time when the load is on the other side of the repressor whose concentration is increasing. “Both Sides Rise” or decay refer to the rise and decay times when a load is present on both sides (symmetrically). The plot shows that at every Kd, the relation between the response time and load is approximately linear. The response time is largest for the case of “Both Sides Rise” followed by “Opposite Side Rise”. The response time is also non-monotonic with respect to the Kd for a given load, and is maximized at intermediate values of Kd.

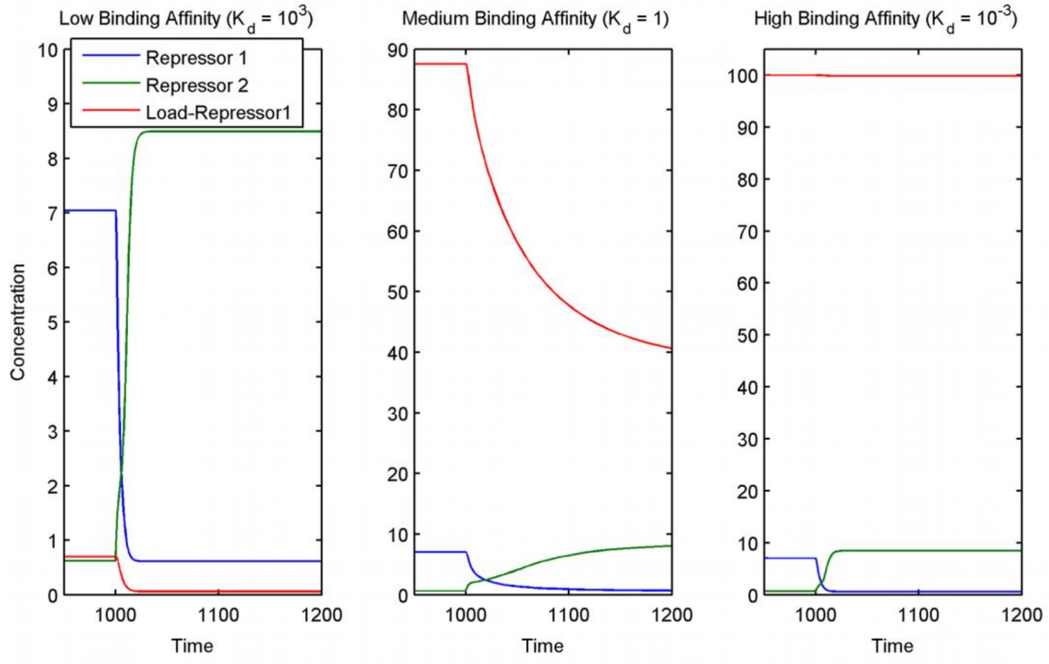


Figure S4.2. Time plot of switching of the simple toggle switch with a load on Repressor 1, at three different values of the dissociation constant.

In all three cases the system is switched by providing 150 molecules/µm3 of an inducer at 1000 minutes. The inducer stays constant at that value and is not shown in the plots. The left panel has a very high dissociation constant (Kd=1000 molecules/µm3) of binding between the load and the repressor, due to which the load has a minimal effect on the system. The middle panel has an intermediate value (Kd=1 molecules/µm3) because of which the load acts as a dynamic sink by releasing Repressor 1 and slowing the switching. The right panel shows the effect of a small dissociation constant (Kd=10−3molecules/µm3). At such strong binding affinities, all of the load is always bound to Repressor 1. Thus the load has minimal effect on the switching dynamics. In all cases total load concentration is 100 molecules/µm3.

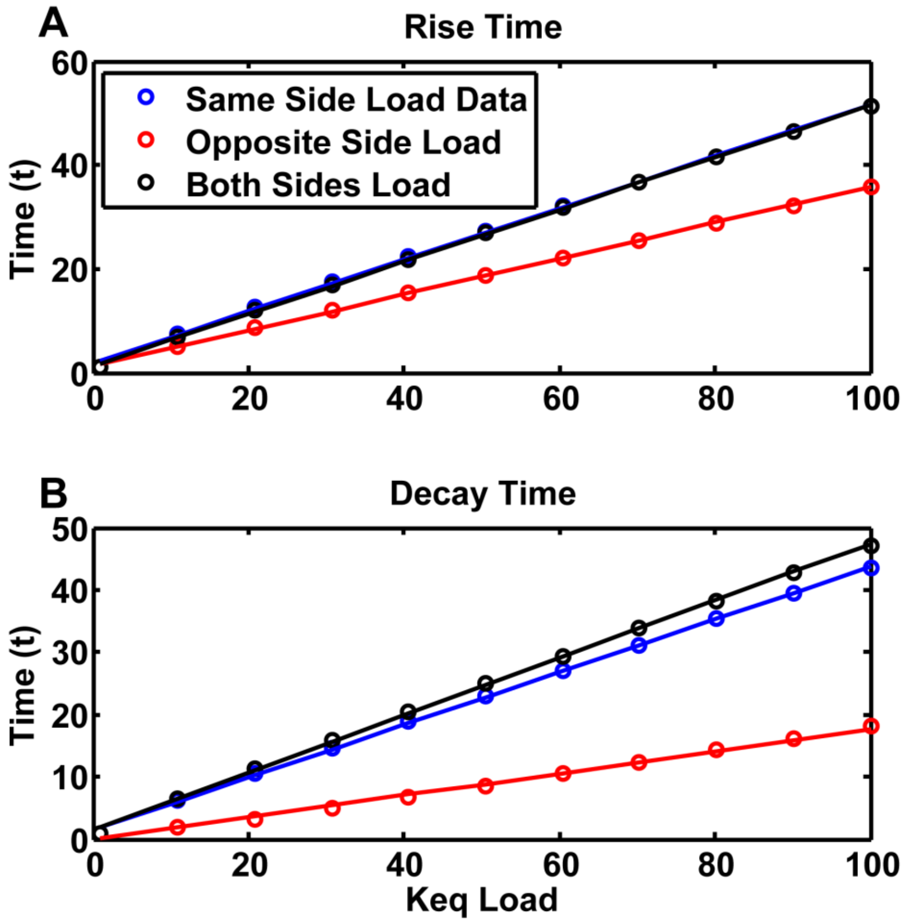


Figure S4.3. Effects of a dynamic load on dynamics of a symmetric toggle switch.
(A). The time taken to reach 90% of maximum value for the protein undergoing a low-to-high transition as a function of the equilibrium constant of a dynamic load. Normalized time is a unit-less number defined by the transition time (rise or decay) of the system at a given loading condition divided by the transition time (rise or decay) of an unloaded system. (B). The time taken for the concentration of the protein undergoing a high-to-low transition to reach 10% of its maximum value.

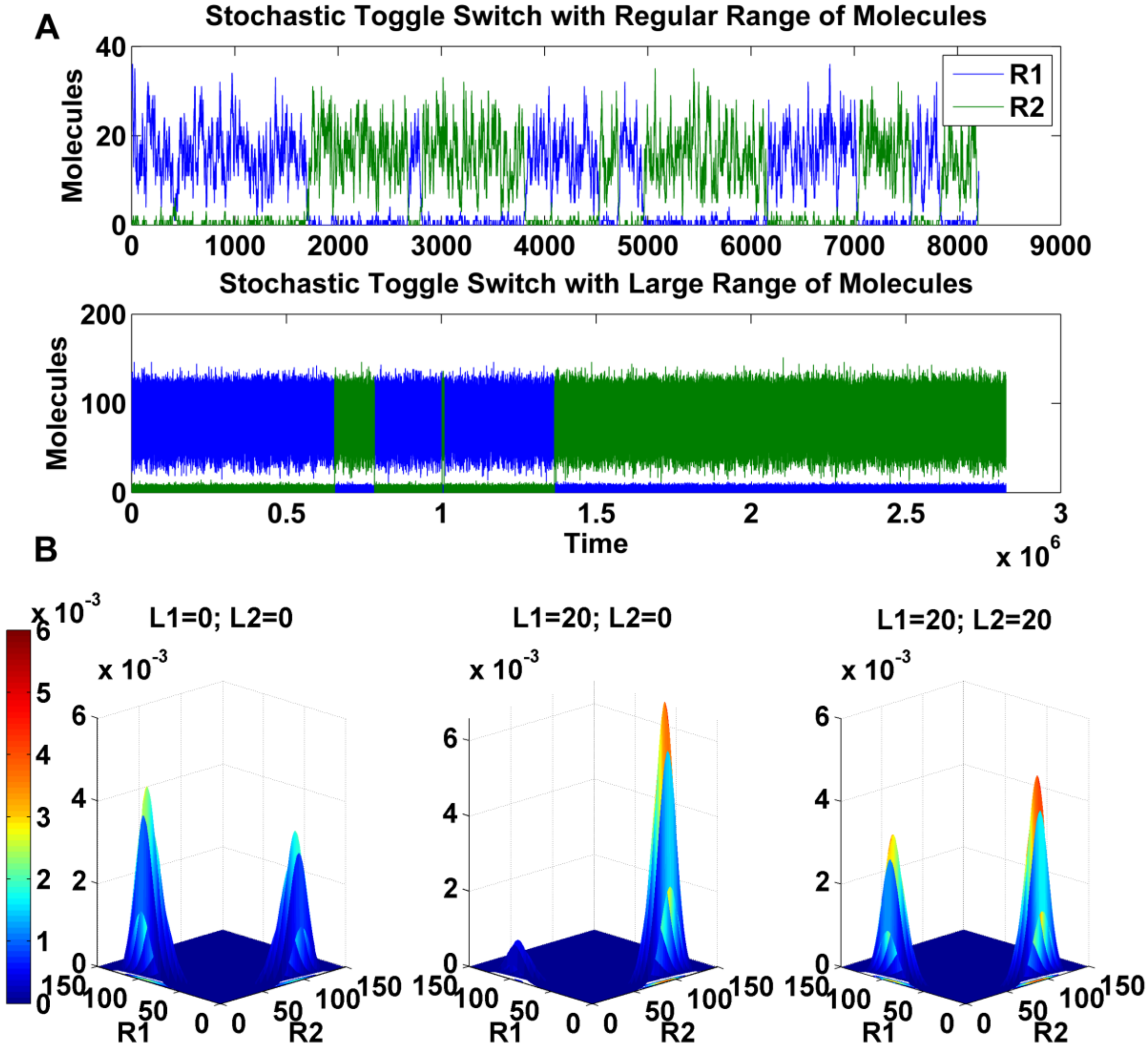


Figure S4.4. Stochastic time trace and the probability distribution function of repressor concentrations for the large volume simulations.

(A). Comparison of time traces of the stochastic simulations of the simple toggle switch with basal parameters (top panel) and a larger volume (bottom panel). The average molecule number is about 5 times greater, and the number of transitions are significantly fewer. (B). The probability distribution function of the genetic toggle switch with the larger molecular number without (left) and with (right) a load. The effect of a load on R1 is qualitatively the same for this system as for the smaller system. Since transitions are slower the data are more uneven for this simulation.

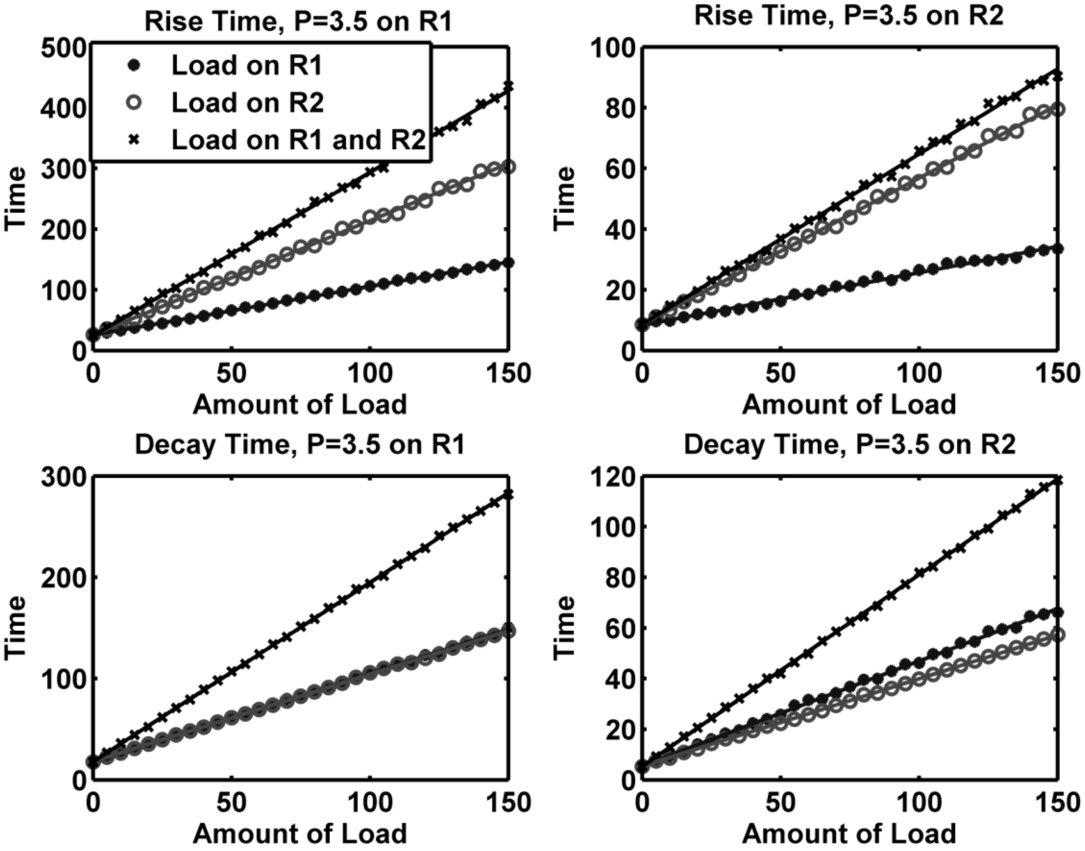


Figure S4.5. Transition times in a genetic toggle switch with a positive feedback moiety. In all cases the strength of the positive feedback (denoted here by P instead of ρ) is 3.5 on either Repressor 1 (R1) or Repressor 2 (R2). Top Left: Rise time - time to transition INTO state R1 with the positive feedback on R1. Note that the rise time is larger at nonzero loads when the load is on R2 or when the load is on both sides, in agreement with the simple toggle switch. Top Right: Rise time - time to transition INTO state R1 with the positive feedback on R2. Bottom Left: Decay time - time to transition OUT OF state R1 with the positive feedback on R1. Bottom Right: Decay time - time to transition OUT OF state R1 with the positive feedback on R2.

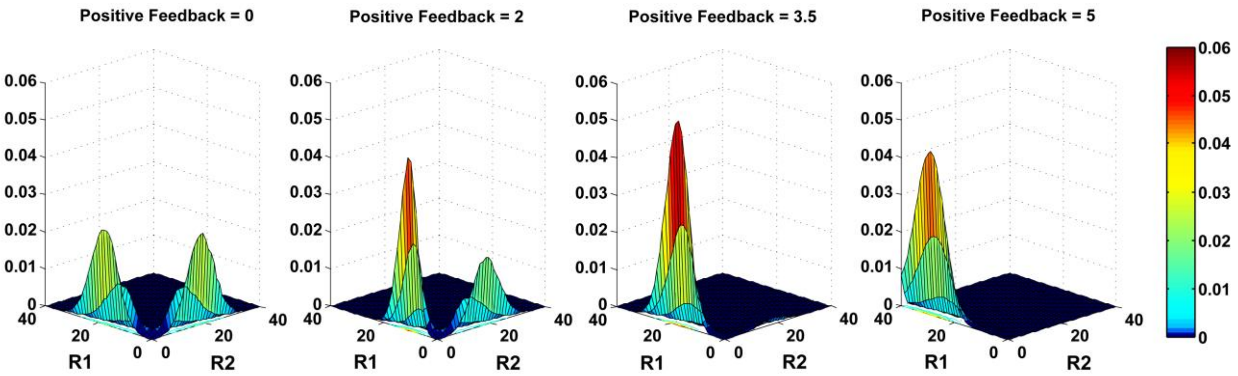


Figure S4.6. Probability distribution functions of repressor concentrations for the toggle with a positive feedback moiety.

The left panel shows that when ρ=0, the switch is balanced evenly. As ρ increases, the side of the switch with the positive feedback becomes more and more prominent, at the expense of the other side. When ρ=5, the system spends most of its time in one state.

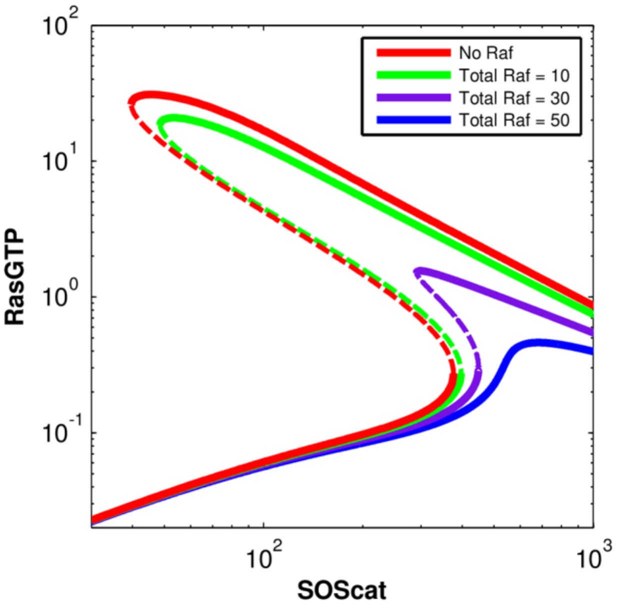


Figure S4.7. Bifurcation diagram of the Ras switch with different levels of Raf (load) on the system for the model with Pseudo Steady State Assumption (PSSA).

The total number of SOS in the simulation box is used as the parameter being tuned, which varies from 0 to 1000. For Raf=0, Raf=10 and Raf=30, there are two bifurcations points as SOS is increased. In the first bifurcation a new high valued stable steady state appears along with the low valued stable steady state. In the second bifurcation, the low valued stable state disappears leaving behind only the high valued state. The dotted line marks the unstable steady state that also comes into existence in the bistable region. As total Raf increases, the two bifurcations approach each other. When Raf=50, the system has lost both of its bifurcations and is characterized by a single stable steady state at all values of Raf.

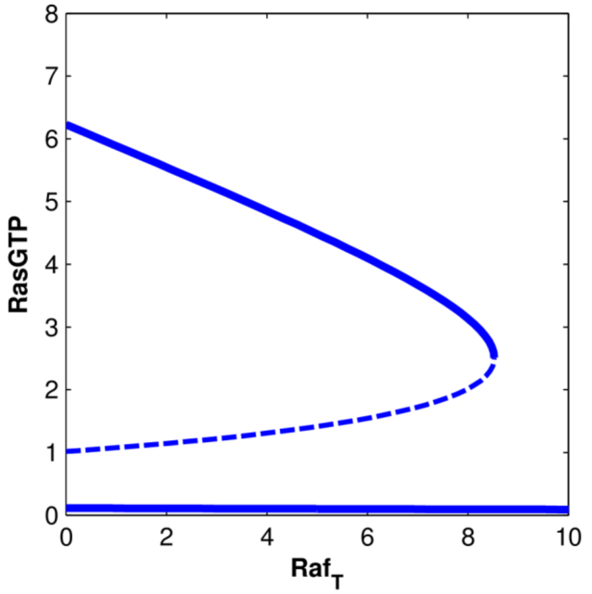


Figure S4.8. Bifurcation diagram of the Ras activation model based on Law of Mass Action (LMA).

Here the total number of Raf molecules (RafT) is the primary parameter being varied. Without Raf, the Ras activation system is bistable as reported. With increasing RafT, the “high” stable steady state branch comes closer with the unstable steady state branch and both are eliminated after a threshold of RafT. A monostable region is maintained beyond the threshold.

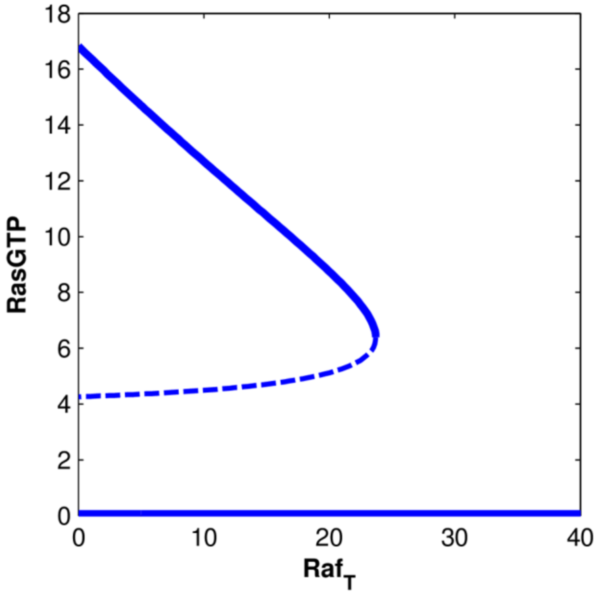


Figure S4.9. Bifurcation diagram of the Ras activation PSSA model with total number of Raf molecules (RafT) as the primary parameter.

Without Raf, the Ras activation system is bistable as reported. With increasing RafT, the “high” stable steady state branch comes closer with the unstable steady state branch and both are eliminated after a threshold of RafT. A monostable region is maintained beyond the threshold.

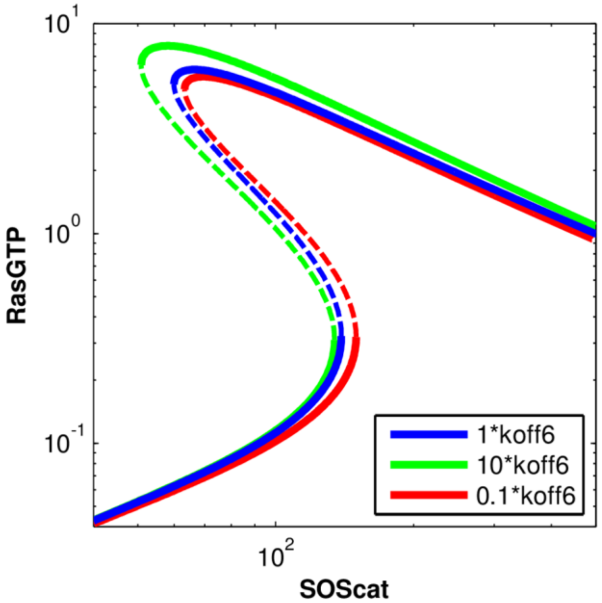


Figure S4.10. Parameter sensitivity of the bistability of Ras switch to changes in koff6. Increase of koff6results in leftward shifts of both stable fixed points, increase in the bistable regime and increase in maximal RasGTP activation level (Green Line) when compared to baseline with original value (Blue Line). Decrease of koff6 (Red Line) results in right shift of both limit points, increase in unstable bistable regime and decrease in maximal RasGTP activation level. Qualitative features of bistability are maintained.

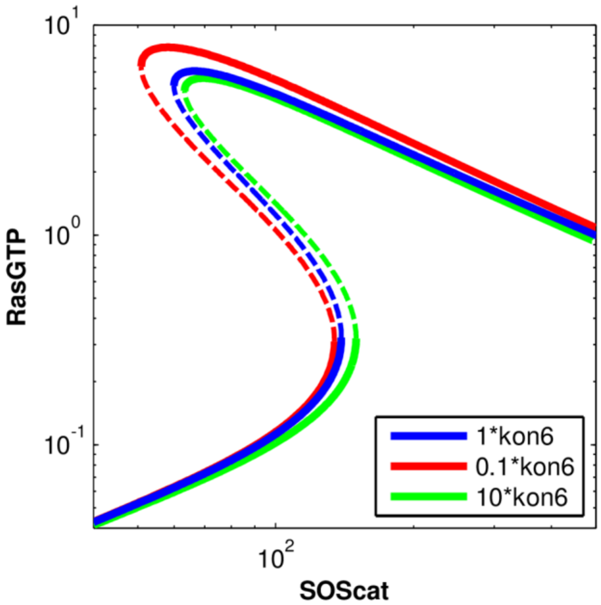


Figure S4.11. Parameter sensitivity of the bistability of Ras switch to changes in kon6. Increase of kon6(Green Line) results in right shift of both limit points, increase in unstable bistable regime and decrease in maximal RasGTP activation level when compared to baseline original value (Blue Line). Decrease of kon6 (Red Line) results in left shifts of both limit points, increase in bistable regime and increase in maximal RasGTP activation level. Qualitative features of bistability are maintained.

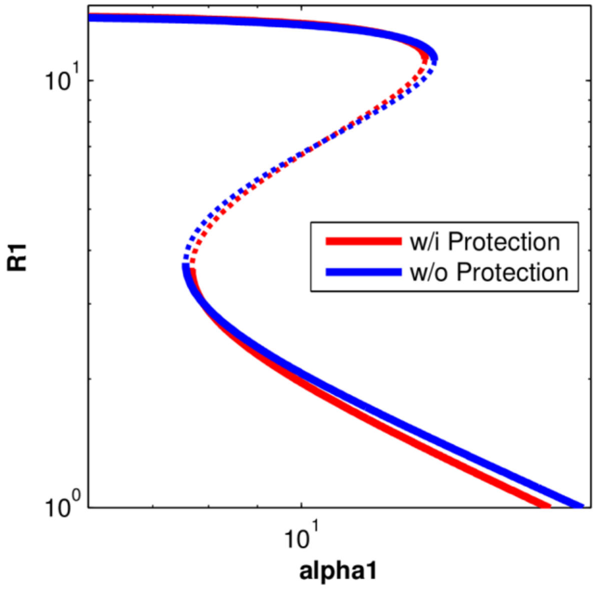


Figure S4.12. Comparison between bifurcation diagrams of toy genetic toggle switch with and without protection of repressor degradation when bound with promoters.

If the protection is not included (Blue Line), a minor increase in the bistable region can be observed with right shift of upper limit point and left shift of lower limit point compared to the case with protection assumed (Red Line). Note that this is not the same as degradation after being bound with the load.

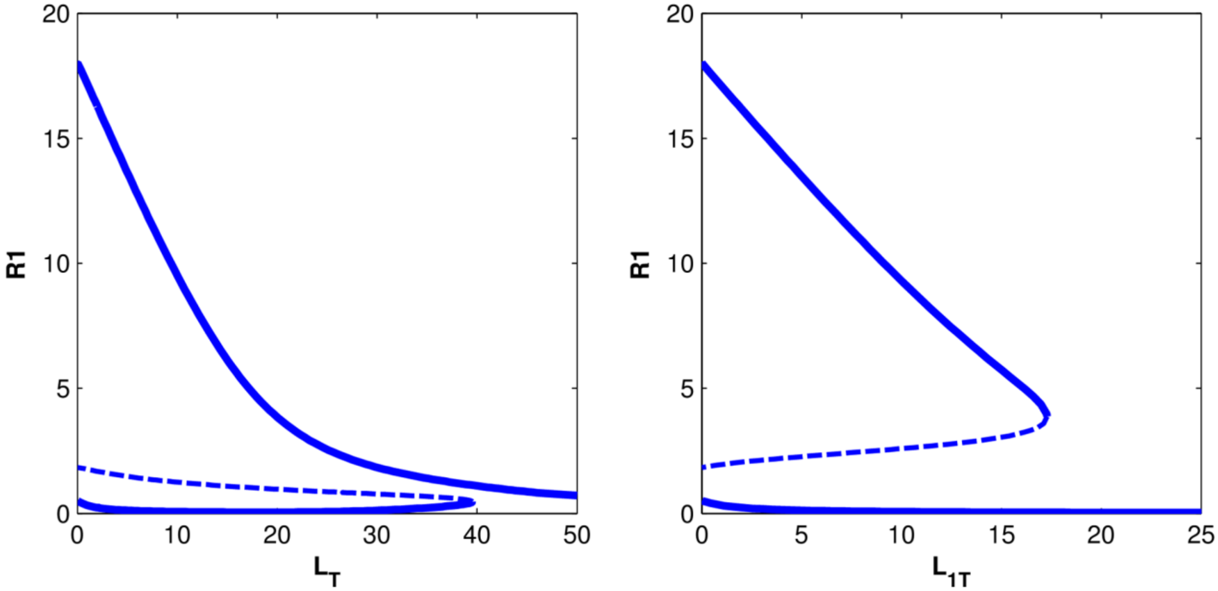


Figure S4.13. Bifurcation diagram of the genetic toggle switch with positive feedback loop on one side after removal of the protection assumption.
The left panel shows the bifurcation diagram when the load is added symmetrically to both sides. Without load molecule, the toggle switch is bistable as predicted. With the increase in LT, the unstable steady state and the "low" stable steady state come closer and meet at certain threshold. The value of "high" stable steady state decreases with increase in LT. Beyond the threshold, the toggle switch becomes monostable. The right panel shows the effect of just adding a load to R1. In this case the high state of R1 approaches the unstable steady state, and annihilates itself. The system jumps to the low stable state, which is equivalent to the "high" state of the other repressor.

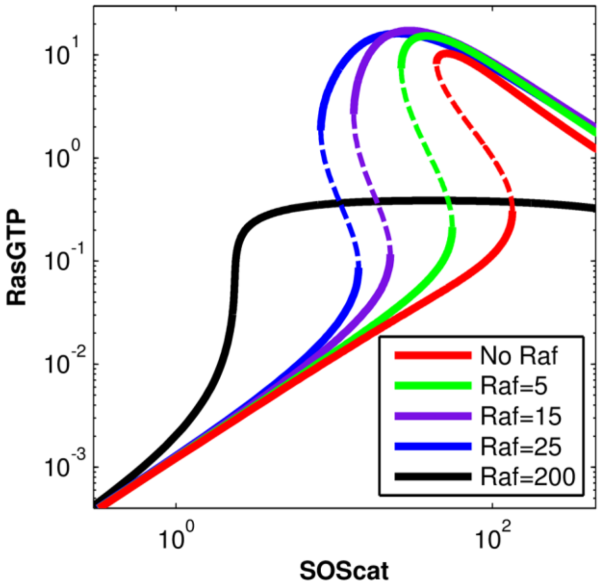


Figure S4.14. Bifurcation diagram of the Ras activation model when Ras can degrade when bound with Raf.

As the number of Raf molecules increase, the bistable region decreases. However unlike the case with no protection, the curve moves to the left. When Raf molecules increase by a large amount, bistability is abrogated.

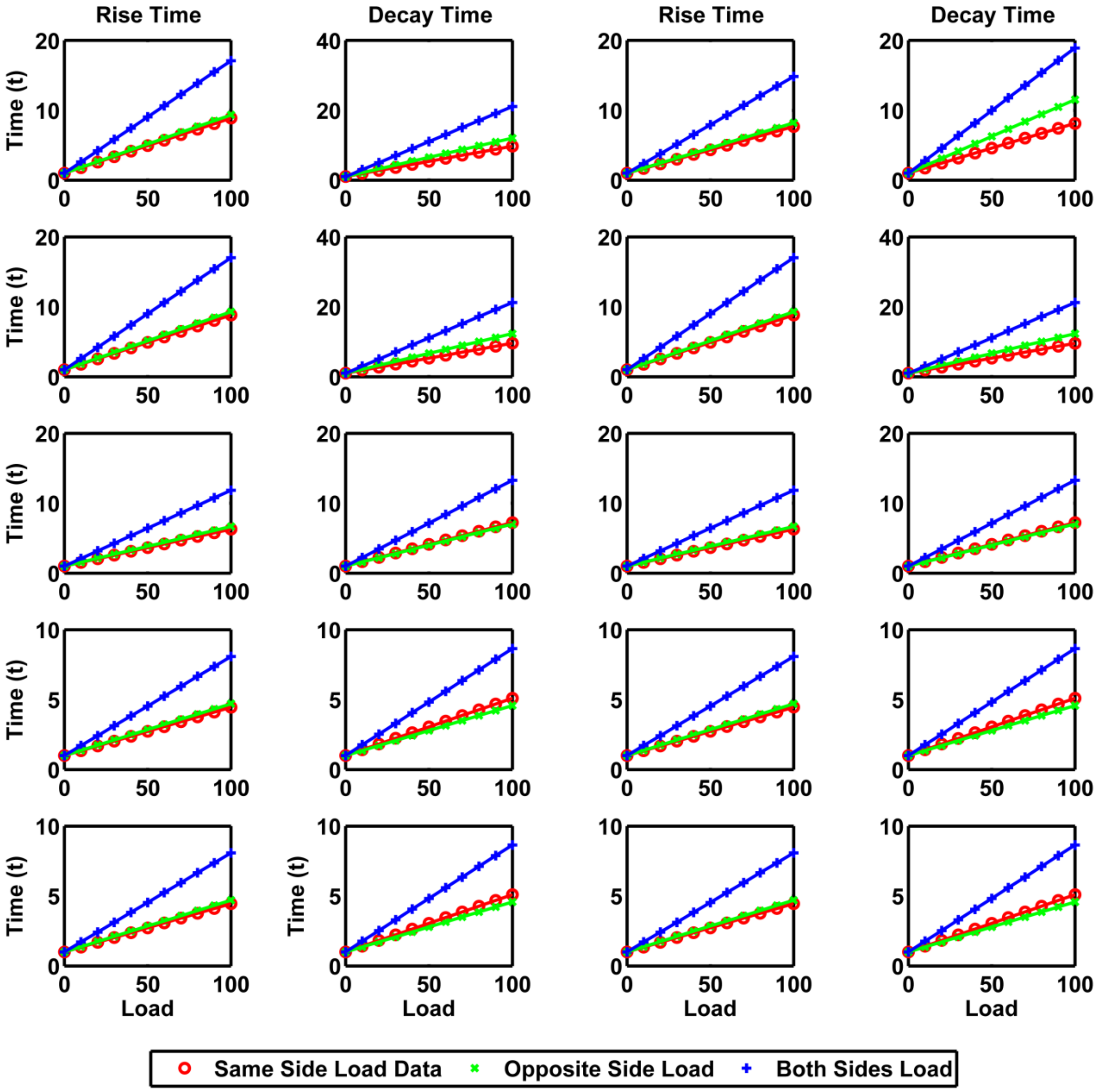


Figure S4.15. Transition times for various k′on and k′off values plotted as a function of load for the basic toggle switch.

Even if the binding-unbinding rates are slower or much faster than protein decay rates, the load-transition time relationship stays linear. A, C, E, G, I, K, M, O, Q and S show the rise time. B, D, F, H, J, L, N, P, R, and T show decay time. (A,B) k′on=4, k′off=0.5, Kd=0.125. (C,D) k′on=10, k′off=0.5, Kd=0.05. (E,F) k′on=4, k′off=4, Kd=1. (G,H) k′on=10, k′off=10, Kd=1. (I,J) k′on=4, k′off=20, Kd=5. (K,L) k′on=10, k′off=50, Kd=5. (M,N) k′on=4, k′off=40, Kd=10. (O,P) k′on=10, k′off=100, Kd=10. (Q,R) k′on=40, k′off=400, Kd=10. (S,T) k′on=100, k′off=1000, Kd=10.

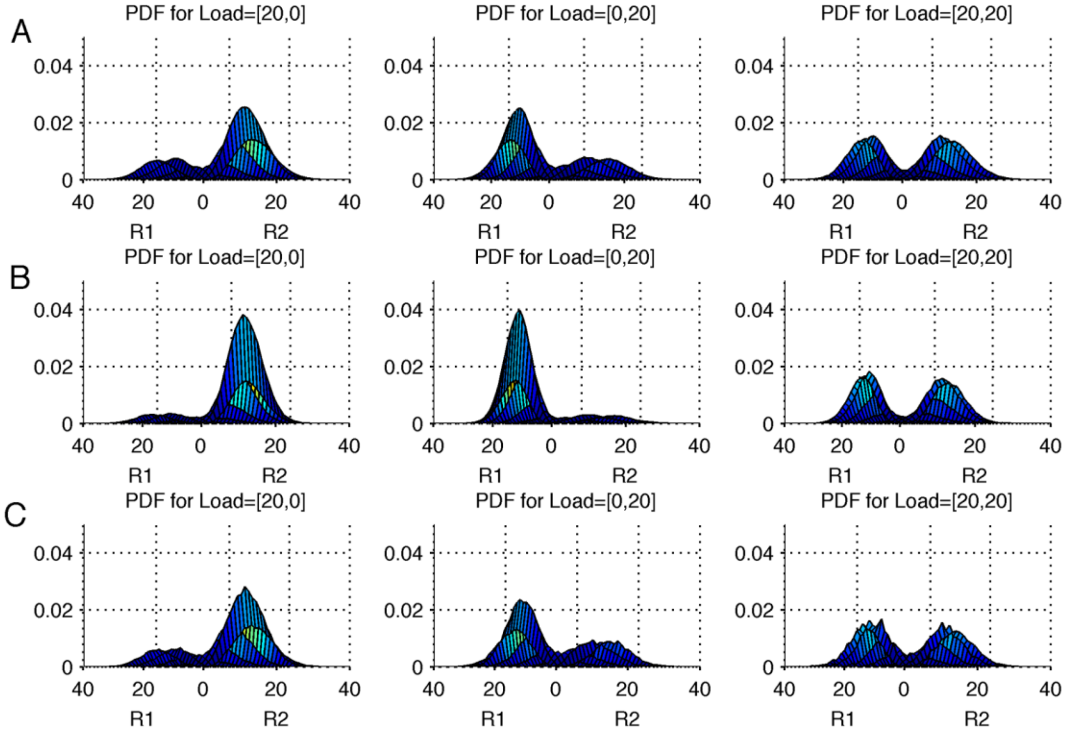


Figure S4.16. Probability distributions of repressor concentrations for various values of k′on and k′offfor the basic toggle switch.

Even when the binding-unbinding with the load is several times faster than protein decay rates, the basic phenomena discussed in the paper remains unchanged. (A) k′on=50, k′off=500 (B) k′on=500 k′off=500 (C) k′on=500 k′off=5000.

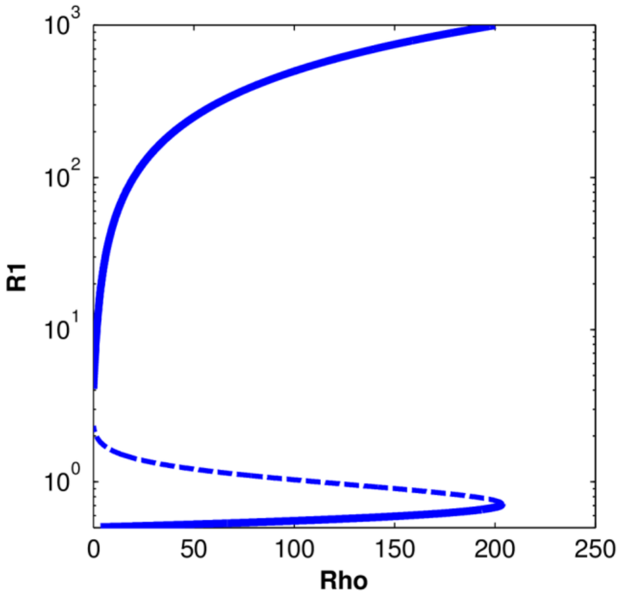


Figure S4.17. Bistability of the toggle switch with positive feedback.
A bifurcation diagram of the simple toggle switch with a positive feedback moiety on one side, with respect to the parameter ρ that measures the strength of the positive feedback. Only the concentration of R1 is shown for simplicity. The switch remains bistable till ρ becomes larger than a little over 200.

8.6 Tables

Table S4.1. Slopes of linear fits to rise and decay time with various values of Koff, Kon and β.

The first column reports the values of the dissociation constant (Kd=Koff/Kon) and the kinetic constants of the binding of Repressor 1, 2 or the value for β, which represents promoter strength. The other columns report the slopes of the linear fits of the various rise times and decay times. In all cases the fits have high R-squared values (>0.95). Intercept is 1, as the slopes are normalized to the un-loaded transition time. For Kd we change the parameters by two orders of magnitude in both directions to show that the linear relation is robust despite these changes. Note that the relation between rise time or decay time and the binding constant is non-monotonic. Units are as reported in the text.

	Rise Time			Decay Time		
	Same-Sided	Opposite Side	Both Sides	Same-Sided	Opposite Side	Both Sides
Koff=0.005; Kon=0.5; Kd= 0.01	4.59E-03	2.12E-03	6.72E-03	1.76E-03	1.01E-02	1.19E-02
Koff=0.05; Kon=0.5; Kd= 0.1	2.46E-02	2.75E-02	5.19E-02	2.50E-02	4.47E-02	7.10E-02
Koff=0.5; Kon=0.5; Kd= 1	7.80E-02	8.32E-02	1.60E-01	8.59E-02	1.13E-01	2.03E-01
Koff=5; Kon=0.5; Kd= 10	3.46E-02	3.71E-02	7.08E-02	4.11E-02	3.57E-02	7.66E-02
Koff=50; Kon=0.5; Kd= 100	4.68E-03	4.98E-03	9.62E-03	5.51E-03	4.25E-03	9.72E-03
Koff=0.5; Kon=50; Kd= 0.01	2.50E-04	3.93E-04	6.44E-04	3.34E-04	4.68E-04	8.01E-04
Koff=0.5; Kon=5; Kd= 0.1	2.83E-03	3.31E-03	6.14E-03	3.05E-03	5.36E-03	8.41E-03
Koff=0.5; Kon=0.5; Kd= 1	7.80E-02	8.32E-02	1.60E-01	8.59E-02	1.13E-01	2.03E-01
Koff=0.5; Kon=0.05; Kd= 10	3.45E-02	3.80E-02	7.15E-02	4.28E-02	3.47E-02	7.77E-02
Koff=0.5; Kon=0.005; Kd= 100	5.29E-03	5.24E-03	1.07E-02	5.89E-03	4.12E-03	1.03E-02
β1= β2 = 0.4	7.22E-02	1.48E-02	1.42E-01	2.08E-01	3.76E-02	2.46E-01
β1= β2 = 4	3.69E-03	1.62E-01	1.50E-01	5.29E-03	8.70E-02	9.76E-02
β1= β2 = 40	-	2.09E-01	2.02E-01	3.68E-05	9.20E-02	9.23E-02

Table S4.2. Exponential Fits of the amount of inducer required to transition states as a function of load.

The basic genetic toggle switch switch was toggled to its other state by production of the other repressor protein by an inducer, given here as a bolus with a decay rate as shown. The size of the bolus was increased until the state changed. This was repeated at different levels of load and the minimum size of the bolus required was fit by an exponential function of the load. The fits are shown here, along with their R-squared values. “Load applied to the opposite side” means switching from a state without a load to a state with a load. “Load applied to the same side” means switching from a state with a load to a state without a load.

Inducer Decay Rate (1/min)	Equation	R^2 Value
	Load Applied to both sides	
0.5	Inducer = 18.44*exp(0.305*Load)	0.99971
0.1	Inducer = 2.41*exp(0.0629*Load)	0.99627
0.05	Inducer = 2.00*exp(0.0308*Load)	0.99995
0.01	Inducer = 1.69*exp(0.006147*Load)	0.99271
0.005	Inducer = 2.23*exp(0.00294*Load)	0.99204
	Load Applied to the “opposite side”	
0.5	Inducer = 16.80*exp(0.108*Load)	0.999188
0.1	Inducer = 1.94*exp(0.0224*Load)	0.994335
0.05	Inducer = 1.54*exp(0.0111*Load)	0.993154
0.01	Inducer = 1.05*exp(0.00233*Load)	0.998114
0.005	Inducer = 0.972*exp(0.00116*Load)	0.996343
	Load Applied to the “same side”	
0.5	Inducer = 17.7*exp(0.219*Load)	0.999907
0.1	Inducer = 2.34*exp(0.0484*Load)	0.996021
0.05	Inducer = 1.91*exp(0.0239*Load)	0.997584
0.01	Inducer = 1.45*exp(0.00496*Load)	0.992406
0.005	Inducer = 1.46*exp(0.00245*Load)	0.991991

Table S4.3. Exponential Fits of the amount of inducer required to transition states as a function of load, in the case of induction by repression.

The switch was toggled to its other state by repression of the current state by an external molecule, given to the system as a bolus with a decay rate as shown. The size of the bolus was increased until the state changed. This was repeated at different levels of load and the minimum size of the bolus required was fit by an exponential function of the load. The fits are shown here, along with their R-squared value. Thus the inducer required depends exponentially on the load in both the methods of induction. “Load applied to the opposite side” means switching from a state without a load to a state with a load. “Load applied to the same side” means switching from a state with a load to a state without a load.

Inducer Decay Rate	Equation	R^2 Value
	Load Applied to Both Sides	
0.5	Inducer = 43.56*exp(0.506*Load)	0.999911
0.1	Inducer = 5.61*exp(0.111*Load)	0.998743
0.05	Inducer = 3.90*exp(0.0586*Load)	0.993461
0.01	Inducer = 3.08*exp(0.0117*Load)	0.991054
0.005	Inducer = 3.07*exp(0.00580*Load)	0.991298
	Load Applied to Opposite Side	
0.5	Inducer = 45.13*exp(0.0700 *Load)	0.999526
0.1	Inducer = 5.12*exp(0.0127*Load)	0.999407
0.05	Inducer = 3.42*exp(0.00663*Load)	0.996845
0.01	Inducer = 2.68*exp(0.00130*Load)	0.99715
0.005	Inducer = 2.49*exp(0.000665*Load)	0.994754
	Load Applied to Same Side	
0.5	Inducer = 47.01*exp(0.413*Load)	0.999474
0.1	Inducer = 8.35*exp(0.0839*Load)	0.993501
0.05	Inducer = 5.35*exp(0.0450*Load)	0.995209
0.01	Inducer = 4.98*exp(0.00881*Load)	0.995697
0.005	Inducer = 4.54*exp(0.00450*Load)	0.993970

Table S4.4. Slopes of linear fits to rise and decay time with a dynamic load, with varying values of load decay rate Kd, load binding rates Kon and Koff, and constant K1.
The first four columns report the values of the various parameters. The other columns report the slopes of the linear fits of the various rise times and decay times. In most cases the fits have high R-squared values (>0.95). The two exceptions are >0.90 and starred. Intercept is 1, as the slopes are normalized to the un-loaded transition time. Note that for all cases, the relationship between load (expressed here as Keq=Kb/Kd) and transition time is a positive linear relationship.

				Rise Time			Decay Time		
Kd	K1	Kon	Koff	Same	Opposite	Both	Same	Opposite	Both
0.5	1	0.5	0.5	0.498	0.345	0.502	0.423	0.179	0.460
0.5	**5**	0.5	0.5	1.333	0.518	0.340	0.463	0.773	0.435
0.5	**0.5**	0.5	0.5	0.264	0.224	0.343	0.308	0.089	0.337
0.05	1	0.5	0.5	0.044	0.483	0.547	0.622	0.058	0.628
5	1	0.5	0.5	0.357	0.332	0.437	0.393	0.187	0.427
0.5	1	**5**	0.5	4.367	0.540	0.520	0.535	1.767	0.578
0.5	1	**0.05**	0.5	0.062	0.059	0.100	0.093	0.014	0.105
0.5	1	0.5	**5**	0.057	0.055	0.093	0.084	0.017	0.093
0.5	1	0.5	**0.05**	4.538	0.582	0.506	0.578	1.835	0.625

Table S4.5. Rate expressions used for the stochastic simulations of the genetic toggle switch.
The rate expressions used for the stochastic simulation of the toggle switch along with the description of the reaction are listed.

Rxn	Rate Expression	h_i	Description of rate
1	h_1*R1	α_1*V	Basal production promoter 1
2	h_2*R1	β_1/(1+R2/V)^n_1)	Repressed production promoter 1
3	-h_3*R1	D*R1	Degradation
4	h_4*R1-h_4[R1:L1]	k_{off}*[R1:L1]	Unbinding from load
5	-h_5*R1+h_5[R1:L1]	k_{on}*R1	Binding to load
6	h_6*R2	α_2*V	Basal production promoter 2
7	h_7*R2	β_2/(1+R2/V)^n_2)	Repressed production promoter 2
8	-h_8*R2	D*R2	Degradation
9	h_9*R2-h_9[R2:L2]	k_{off2}*[R2:L2]	Unbinding from load
10	-h_{10}*R2+h_{10}[R2:L2]	k_{on2}*R2	Binding to load

Table S4.6. List of reactions in the minimal model of Ras activation.
The reactions in the minimal model of Ras activation, along with the labels of the corresponding rate constants are shown. Parameters used in the simulations are given in Table S4.7.

$Reactions$
$SOScat + RasGDP \xleftrightarrow{kon\,1,koff\,1} SOScat(RasGDP)$
$SOScat + RasGTP \xleftrightarrow{kon\,2,koff\,2} SOScat(RasGTP)$
$(RasGTP) + RasGDP \xleftrightarrow{kon\,3,koff\,3} SOScat(RasGTP): RasGDP \xrightarrow{kcat\,3} SOScat(RasG$ $+ RasGTP$
$RasGDP) + RasGDP \xleftrightarrow{kon\,4,koff\,4} SOScat(RasGDP): RasGDP \xrightarrow{kcat\,4} SOScat(RasG$ $+ RasGTP$
$RasGAP + RasGTP \xleftrightarrow{kon\,5,koff\,5} RasGAP: RasGTP \xrightarrow{kcat\,5} RasGAP + RasGDP$
$RasGTP + Raf \xleftrightarrow{kon\,6,koff\,6} RasGTP: Raf$

Table S4.7. Kinetic rate parameters used for the simulations of the Ras model. Here the numbers in the subscript of the rate constants in the “Constant” column refer to the reactions shown in the corresponding row of Supplementary Table S4.6. The meaning of the rate constants are as follows: kon refers to the on-rate, koff is the off rate and kcat is the catalytic rate. The sources for the rates are as shown in the last column.

Rxn	Constant	3D Rate Values	Units	2D Rate Values	Units	Reference
1	k_{on1}	0.12	μM^{-1} s^{-1}	0.028	Molecules^{-1} s^{-1}	[2]
1	k_{off1}	3.0	s^{-1}	3.0	s^{-1}	[2]
2	k_{on2}	0.11	μM^{-1} s^{-1}	0.026	Molecules^{-1} s^{-1}	[2]
2	k_{off2}	0.4	s^{-1}	0.4	s^{-1}	[2]
3	k_{on3}	0.05	μM^{-1} s^{-1}	0.0118	Molecules^{-1} s^{-1}	[2]
3	k_{off3}	0.1	s^{-1}	0.1	s^{-1}	[2]
3	k_{cat3}	0.038	s^{-1}	0.038	s^{-1}	[2]
4	k_{on4}	0.07	μM^{-1} s^{-1}	0.0165	Molecules^{-1} s^{-1}	[2]
4	k_{off4}	1.0	s^{-1}	1.0	s^{-1}	[2]
4	k_{cat4}	0.003	s^{-1}	0.003	s^{-1}	[2]
5	k_{on5}	1.74	μM^{-1} s^{-1}	0.41	Molecules^{-1} s^{-1}	[2]
5	k_{off5}	0.2	s^{-1}	0.2	s^{-1}	[2]
5	k_{cat5}	0.1	s^{-1}	0.1	s^{-1}	[2]
6	k_{on6}	29.6e6	M^{-1} s^{-1}	6.96	Molecules^{-1} s^{-1}	[3]
6	k_{off6}	5.22	s^{-1}	5.22	s^{-1}	[3]

Table S4.8. List of reactions in the toy model of genetic toggle switch.
The reactions in the toy model of the genetic toggle switch, discussed in Supplementary Text S4.1 are listed. The description of the various chemical species in the reactions are also provided in the Supplementary Text S4.1.

Reactions	*Index*
AR_1 *Module*	
$R_1 + R_1 + R_1 \xrightleftharpoons[k_2]{k_1} AR_1$	*P1*
$AR_1 + Pro_1 \xrightleftharpoons[k_4]{k_3} AR_1 : Pro_1$	*P2*
$Pro_1 \xrightarrow{a_2} Pro_1 + R_2$	*P3*
$AR_1 : Pro_1 \xrightarrow{k_5} Pro_1$	*P4*
$AR_1 \xrightarrow{k_6} \varnothing$	*P5*
$R_1 \xrightarrow{k_7} \varnothing$	*P6*
$R_1 + L_1 \xrightleftharpoons[k_{off1}]{k_{on1}} R_1 : L_1$	*P7*
AR_2 *Module*	
$R_2 + R_2 + R_2 \xrightleftharpoons[k_9]{k_8} AR_2$	*P8*
$AR_2 + Pro_2 \xrightleftharpoons[k_{11}]{k_{10}} AR_2 : Pro_2$	*P9*
$Pro_2 \xrightarrow{a_1} Pro_2 + R_1$	*P10*
$AR_2 : Pro_2 \xrightarrow{k_{12}} Pro_2$	*P11*
$AR_2 \xrightarrow{k_{13}} \varnothing$	*P12*
$R_2 \xrightarrow{k_{14}} \varnothing$	*P13*
$R_2 + L_2 \xrightleftharpoons[k_{off2}]{k_{on2}} R_2 : L_2$	*P14*

REFERENCES

[1] Das J, Ho M, Zikherman J, Govern C, Yang M, et al. (2009) Digital signaling and hysteresis characterize ras activation in lymphoid cells. Cell 136: 337-351.

[2] Prasad A, Zikherman J, Das J, Roose JP, Weiss A, et al. (2009) Origin of the sharp boundary that discriminates positive and negative selection of thymocytes. Proc Natl Acad Sci U S A 106: 528-533.

[3] Kiel C, Serrano L (2009) Cell type-specific importance of ras-c-raf complex association rate constants for MAPK signaling. Science Signaling 2: ra38.

APPENDIX III: SUPPLEMENTARY INFORMATION FOR CHAPTER 5

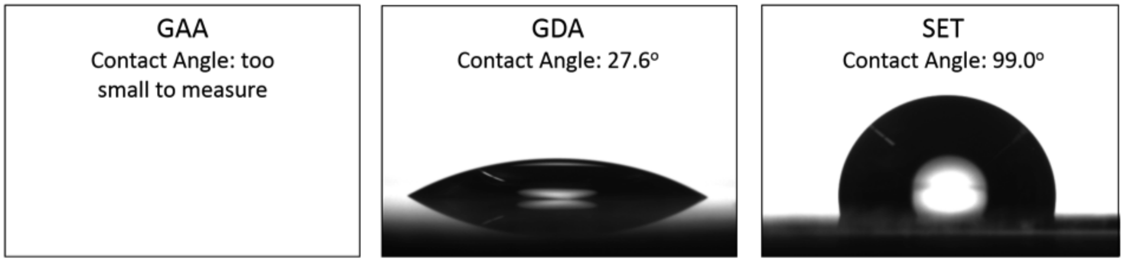


Figure S5.1: Contact angle measurements of the surfaces.
Side view of a drop on GAA, GDA, and SET substrates. Hydrophobicity increases from left to right. Contact angle was too small to be measured for GAA substrates; 27.6° on average for GDA substrates, and 99.0° on average for SET.

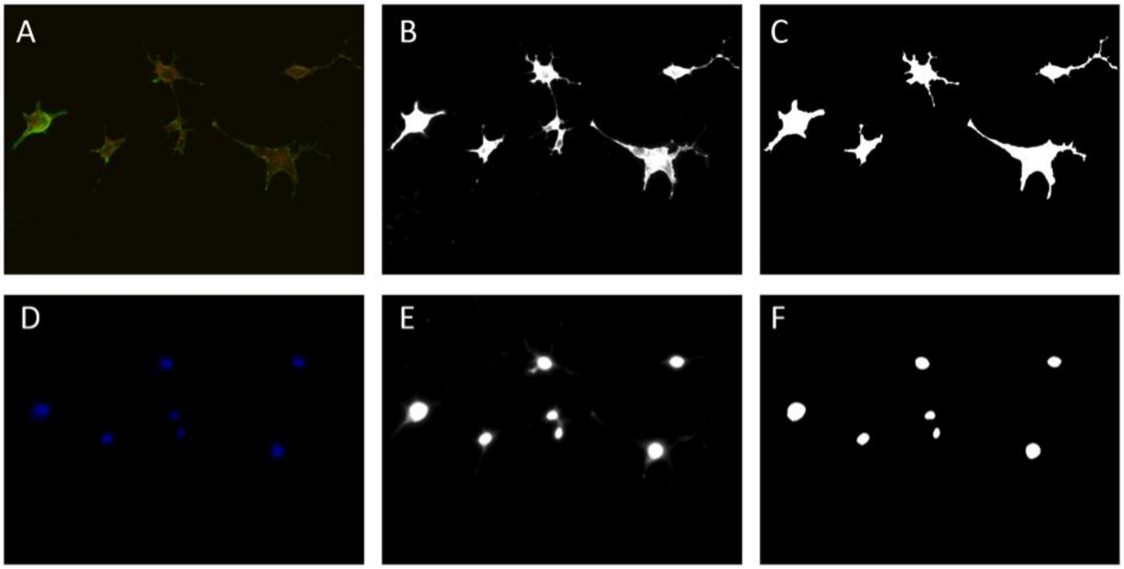


Figure S5.2: Image processing example for shape metrics.
A) Membrane and actin images are enhanced and combined as cell image. D) Enhanced nuclei image. B) Sharpened edge boundaries for cell image. E) Sharpened edge boundaries for nuclei image. C) Binary image of cell. F) Binary image of nuclei.

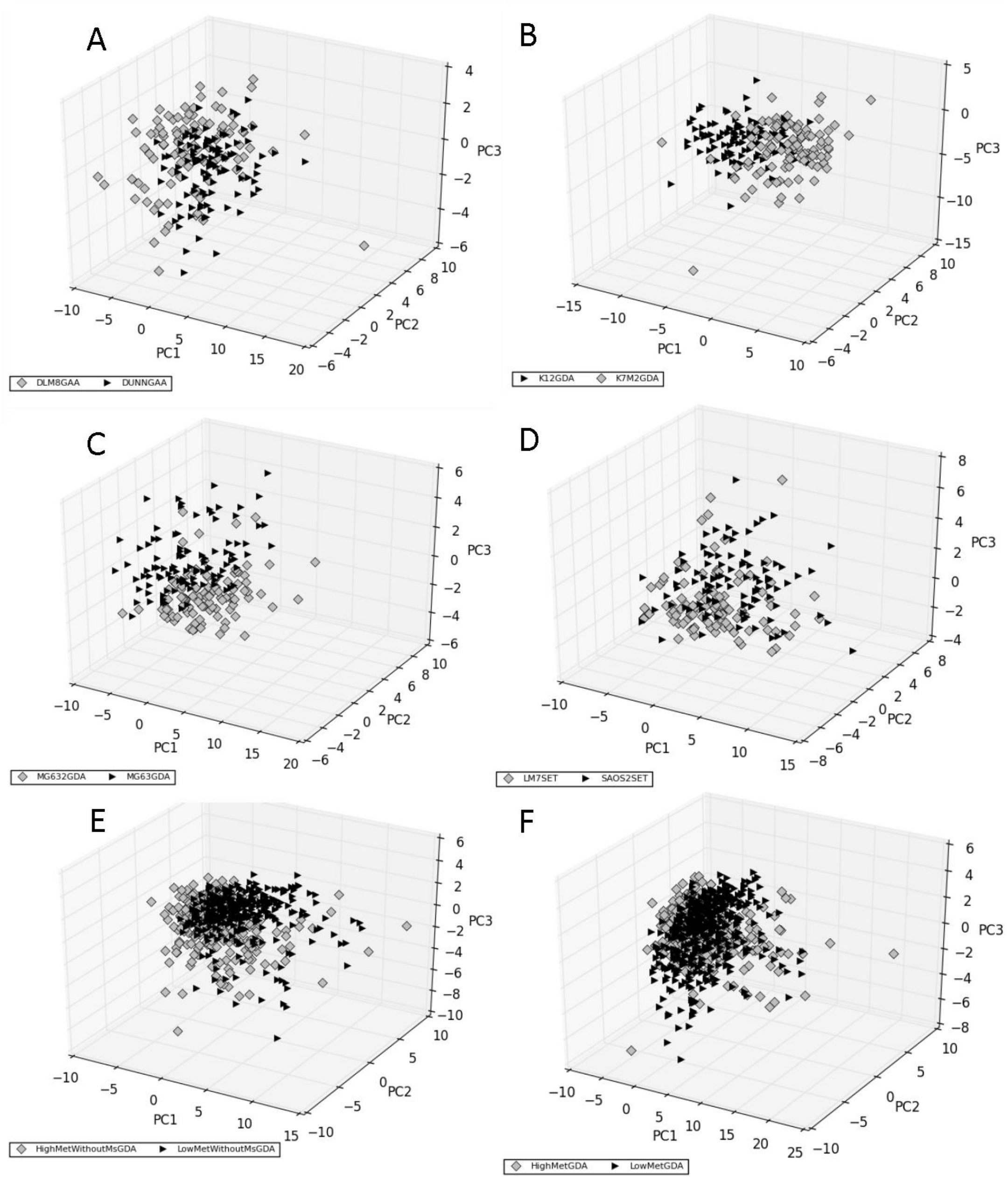


Figure S5.3: Principal component analysis.

The principal component based comparisons that were performed, apart from those reported in the main paper figures. Each panel represents shape data projected on the first three Principal Components of all the morphometric characteristics of the cells. Each panel represents one comparison indicated in the legend. The gray diamond signs represent the high-met lines and the black triangle signs represent the low met lines.

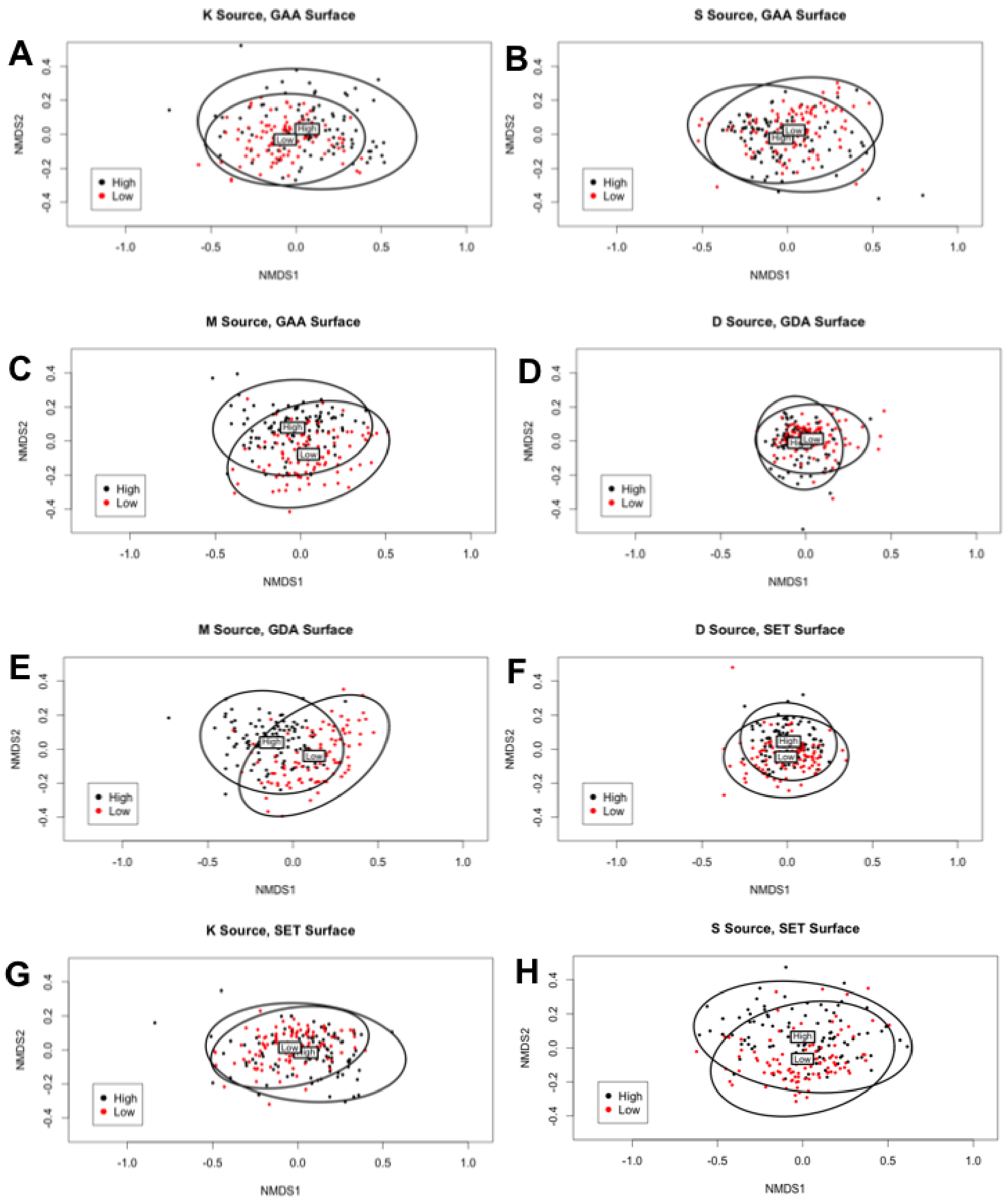


Figure S5.4: Nonmetric multidimensional scaling (NMDS) analysis.
Each panel reports the NMDS based ordination plot of all comparisons we made except for those shown in the main paper figures. Each panel represents a specific comparison, with the legend indicating the pair of cell lines considered and the surface on which they were cultured. The labels "High" and "Low" refer to the data centroids of the high metastatic and low metastatic cell line of the pair. The ellipses represent 95% confidence intervals.

Table S5.1: Morphometric parameters. Description and units of the morphometric measures used for this analysis.

Parameter	Description	Unit
Cell Area	Area in pixels of the cell	pixels^2
Cell Perimeter	Perimeter in pixels of the cell	pixels
Cell Major	The major axis of an ellipse drawn around the cell	pixels
Cell Minor	The minor axis of an ellipse drawn around the cell	pixels
Circularity of Cell	$4\ \pi(\text{Area})/\text{Perimeter}^2$	unitless (0-1)
Aspect Ratio of cell	Major axis of ellipse/minor axis of ellipse	unitless
Roundness of cell	$4(\text{Area})/(\ \pi\ (\text{Elliptical Major Axis})^2)$	unitless (0-1)
Solidity Cell	Area of convex hull of cell / area of the cell	unitless (0-1)
Max span across the hull	The maximum distance from one point on the convex hull to another	pixels
Area of the hull	Area of the convex hull (smallest convex polygon that encloses the cell)	pixels^2
Perimeter of the hull	Perimeter of the convex hull	pixels
Circularity of the hull	$4\pi\ (\text{area of hull/hull perimeter}^2)$	unit less (0-1)
Max radius of the hull	Maximum Distance from centroid of Hull to an exterior point on the hull	pixels
Max/Min Radius of hull	Maximum radius of hull / minimum radius of hull (see max radius definition above)	unitless
CV Rad Hull	The relative variation of radii drawn from the hull's center to an exterior point. Given by the Standard of Deviation of all Radii divided into the mean of all radii	unitless

Mean Radius of hull	The mean of all radii drawn from the hull's centroid to an exterior point	pixels
Diameter of bounding circle	The diameter of the bounding circle drawn around the cell	pixels
Max radius from circle to hull	The maximum distance from the center of the circle to an edge of the convex hull	pixels
Max/Min Radius from Circle	Maximum radius from the center of the circle to an edge of the convex hull / minimum radius from the center of the circle to an edge of the convex hull	unit-less
CV Radius from circle to hull	The relative variation of radii drawn from the circle's center to the hull. Given By the Standard of Deviation of all Radii divided into the mean of all radii	unit-less
Mean radius from circle to hull	The mean of all radii drawn from the circle's center to the hull	pixels
Nucleus area	area of the nucleus in pixels^2	pixels^2
Nucleus perimeter	perimeter of the nucleus in pixels	pixels
Nucleus major	Major axis of an ellipse drawn around the nucleus	pixels
Nucleus minor	Minor axis of an ellipse drawn around the nucleus	pixels
Nucleus circularity	4π (Area)/Perimeter^2	unitless (0-1)
Nucleus aspect ratio	Major axis of ellipse/minor axis of ellipse	unitless
Nucleus roundness	4(Area)/(π (Elliptical Major Axis)^2)	unitless (0-1)
Nucleus solidity	Area of convex hull of cell / area of the cell	unit less (0-1)

Table S5.2. Complete t-test results. The t-test results for paired comparisons between the high met and low met cell line for 29 morphometric characteristics for each paired line on all surfaces. The abbreviations used are as follows: GDA: Glass Detergent washed; GAA: Glass Acid etched; SET: Siliconized glass, Ethanol Treated. L: low metastatic cell line; H: high metastatic cell line. The cell lines are referred to as follows: (i) DL: DUNN; (ii) DH: DLM8; (iii) KL: K12; (iv) KH: K7M2; (v) SL: SAOS2; (vi) SH: LM7-Saos; (vii) ML: MG63; (viii) MH: MG63.2

CELL LINES	**DL vs DH on GDA**			**KL vs KH on GDA**		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	1.70E-07	L	1	9.81E-23	L
Perimeter	1	0.001571742	L	1	1.33E-23	L
Major axis	1	1.00376E-04	L	1	5.13E-09	L
Minor axis	1	7.65E-05	L	1	3.13E-23	L
Circularity	0	0.983677269	H	1	1.35E-13	H
Aspect Ratio	0	0.434960439	H	1	4.90416E-04	H
Roundness	0	0.054961952	L	0	0.734672912	L
Solidity	1	0.019938295	L	1	1.13E-05	H
Max Span Hull	1	9.54706E-04	L	1	7.66E-10	L
Area Hull	1	3.76E-05	L	1	8.95E-18	L
Perimeter Hull	1	1.12529E-04	L	1	3.56E-15	L
Circularity Hull	0	0.111165111	L	1	6.27E-08	L
Max Rad Hull	1	0.00304524	L	1	8.59E-09	L
Max/Min Rad Hull	0	0.451931518	H	1	7.15E-08	H
CV Rad Hull	0	0.345243155	H	1	4.08E-08	H
Mean Rad Hull	1	9.79594E-04	L	1	4.61E-12	L
Diameter Bounding Circle	1	0.001415995	L	1	4.44E-10	L
Max Rad Circle	1	0.001427271	L	1	4.42E-10	L
Max/Min Circle	0	0.311128332	H	1	0.00196701	H
CV Rad Circle	0	0.696827332	H	1	0.001900904	H
Mean Rad Circle	1	0.001463127	L	1	4.57E-11	L
Nucleus Area	1	6.38E-07	L	1	0.040689462	H
Nucleus Perimeter	1	2.33E-07	L	1	0.013025999	H
Nucleus Major	1	2.76E-12	L	1	0.016036836	H
Nucleus Angle	0	0.799829655	L	0	0.750885339	H
Nucleus Circularity	1	0.018320778	H	1	0.004260063	L
Nucleus Aspect Ratio	1	5.11E-07	L	0	0.202738897	H
Nucleus Roundness	1	6.62E-07	H	0	0.077355812	L
Nucleus Solidity	0	0.223859205	H	0	0.227590653	L

	SL v SH on GDA			**ML vs MH on GDA**		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	0.024815142	L	1	3.70E-15	H
Perimeter	1	8.07E-10	L	1	1.68E-20	H
Major axis	1	0.022839291	L	1	1.69E-10	H
Minor axis	1	0.009413983	L	1	3.75E-20	H
Circularity	1	1.94E-18	H	1	1.22E-07	L
Aspect Ratio	0	0.594491625	0	1	0.002108088	L
Roundness	0	0.542498068	0	0	0.222373793	H
Solidity	1	6.67E-10	H	1	1.45E-19	L
Max Span Hull	1	0.001166183	L	1	1.59E-14	H
Area Hull	1	0.002664101	L	1	3.97E-15	H
Perimeter Hull	1	3.33258E-04	L	1	2.09E-18	H
Circularity Hull	0	0.50473601	0	1	0.024119019	H
Max Rad Hull	1	0.001353442	L	1	7.87E-15	H
Max/Min Rad Hull	0	0.391808481	0	1	4.02681E-04	L
CV Rad Hull	0	0.664590878	0	1	7.20870E-04	L
Mean Rad Hull	1	0.001254166	L	1	5.77E-18	H
Diameter Bounding Circle	1	0.001074952	L	1	1.08E-14	H
Max Rad Circle	1	0.001076357	L	1	1.06E-14	H
Max/Min Circle	0	0.316170215	0	1	0.006924285	L
CV Rad Circle	0	0.489636308	0	1	1.80373E-04	L
Mean Rad Circle	1	0.001289779	L	1	6.70E-18	H
Nucleus Area	1	0.007412111	L	0	0.40331311	H
Nucleus Perimeter	1	0.004672354	L	0	0.236410275	L
Nucleus Major	0	0.109388561	0	1	0.003281577	L
Nucleus Angle	1	8.28021E-04	L	0	0.892957546	H
Nucleus Circularity	0	0.47106146	0	1	4.70E-17	H
Nucleus Aspect Ratio	0	0.112842735	0	1	1.17E-16	L
Nucleus Roundness	0	0.056958286	0	1	3.39E-19	H
Nucleus Solidity	1	0.013676416	H	1	1.55E-08	H
	DL vs DH on GAA			**KL vs KH on GAA**		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with

						higher mean
Area	1	1.58E-11	L	1	6.96E-04	L
Perimeter	1	1.67E-13	L	1	1.90E-07	L
Major axis	1	8.26E-06	L	1	0.042396695	L
Minor axis	1	5.50E-15	L	1	1.39E-06	L
Circularity	1	0.001168607	H	1	2.12E-05	H
Aspect Ratio	1	4.26E-07	H	1	0.016219058	H
Roundness	1	2.12E-06	L	0	0.304275226	H
Solidity	0	0.649834591	L	1	4.05E-05	H
Max Span Hull	1	1.93E-06	L	1	0.011594439	L
Area Hull	1	1.38E-09	L	1	6.25E-05	L
Perimeter Hull	1	1.89E-10	L	1	4.99830E-04	L
Circularity Hull	1	2.49E-07	L	1	1.87591E-04	L
Max Rad Hull	1	1.23E-05	L	1	0.015235148	L
Max/Min Rad Hull	1	3.68E-06	H	1	3.54679E-04	H
CV Rad Hull	1	1.45308E-04	H	1	2.04117E-04	H
Mean Rad Hull	1	9.76E-07	L	1	0.003499789	L
Diameter Bounding Circle	1	1.39E-06	L	1	0.011238879	L
Max Rad Circle	1	1.38E-06	L	1	0.011239805	L
Max/Min Circle	1	6.43029E-04	H	1	0.028692264	H
CV Rad Circle	1	0.011462267	H	1	0.00794263	H
Mean Rad Circle	1	1.54E-06	L	1	0.00489324	L
Nucleus Area	1	1.48E-17	L	0	0.075826142	L
Nucleus Perimeter	1	2.36E-19	L	0	0.538210921	L
Nucleus Major	1	2.02E-19	L	0	0.353844427	L
Nucleus Angle	0	0.258661283	H	1	0.020577172	L
Nucleus Circularity	0	0.360119308	H	1	3.54E-06	L
Nucleus Aspect Ratio	0	0.314006166	L	1	0.005623758	H
Nucleus Roundness	0	0.317166606	H	1	5.27773E-04	L
Nucleus Solidity	0	0.075791875	L	1	6.66529E-04	L
	SL v SH on GAA			**ML vs MH on GAA**		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	0.001031243	L	1.00E-00	1.04E-14	H
Perimeter	1	1.33E-07	L	1	3.98E-11	H
Major axis	0	0.422884024	0	1	2.63E-05	H

Minor axis	1	1.68169E-04	L	1	5.21E-18	H
Circularity	1	5.80E-10	H	0	0.402180022	L
Aspect Ratio	1	0.044329963	H	1	0.002538802	L
Roundness	0	0.735616497	0	1	0.016673375	H
Solidity	1	0.001037781	H	1	1.61E-06	L
Max Span Hull	0	0.070738027	0	1	1.66E-05	H
Area Hull	1	3.46386E-04	L	1	1.02E-13	H
Perimeter Hull	1	0.012240483	L	1	1.73E-09	H
Circularity Hull	0	0.163755254	0	1	3.41571E-04	H
Max Rad Hull	0	0.065151994	0	1	5.60E-06	H
Max/Min Rad Hull	0	0.410407701	0	1	5.68E-05	L
CV Rad Hull	0	0.751428165	0	1	2.17E-06	L
Mean Rad Hull	0	0.094881092	0	1	1.27E-09	H
Diameter Bounding Circle	0	0.071050257	0	1	1.20E-05	H
Max Rad Circle	0	0.07089545	0	1	1.19E-05	H
Max/Min Circle	0	0.415344035	0	1	0.017102895	L
CV Rad Circle	0	0.477629979	0	1	6.50E-08	L
Mean Rad Circle	0	0.090905819	0	1	2.86E-09	H
Nucleus Area	1	0.006289331	L	1	0.006707588	L
Nucleus Perimeter	1	0.001425287	L	1	5.23337E-04	L
Nucleus Major	0	0.847030057	0	1	7.52E-09	L
Nucleus Angle	1	6.92E-06	L	0	0.674316166	H
Nucleus Circularity	1	0.002641323	H	1	5.12E-06	H
Nucleus Aspect Ratio	1	3.23E-07	H	1	8.87E-13	L
Nucleus Roundness	1	1.01E-07	L	1	1.98E-13	H
Nucleus Solidity	1	4.08E-06	H	0	0.960896383	H
	DL vs DH on SET			**KL vs KH on SET**		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	7.07E-08	L	1	0.001359293	L
Perimeter	1	1.06E-02	L	1.00E-00	1.92E-04	L
Major axis	0	0.31028643	L	1	0.013883123	L
Minor axis	1	1.18E-05	L	1	0.003635536	L
Circularity	0	0.084699895	L	1	0.00244293	H
Aspect Ratio	1	0.031597212	H	0	0.795298041	L
Roundness	1	1.07E-08	L	0	0.373415036	H

Solidity	1	4.36E-06	L	0	0.16293527	H
Max Span Hull	0	0.537074144	L	1	0.025277059	L
Area Hull	1	0.004863484	L	1	0.021865475	L
Perimeter Hull	0	0.100952481	L	1	0.008627989	L
Circularity Hull	1	2.86E-06	L	0	0.098235037	L
Max Rad Hull	0	0.76158252	L	0	0.062899103	L
Max/Min Rad Hull	1	0.017263712	H	0	0.343331551	H
CV Rad Hull	1	0.001714675	H	0	0.433784196	H
Mean Rad Hull	0	0.425063751	L	0	0.050041448	L
Diameter Bounding Circle	0	0.541368617	L	1	0.024810174	L
Max Rad Circle	0	0.53999839	L	1	0.024854432	L
Max/Min Circle	0	0.243584829	H	0	0.829351169	H
CV Rad Circle	1	0.029595895	H	0	0.590918861	L
Mean Rad Circle	0	0.447429427	L	0	0.0588553	L
Nucleus Area	1	2.20E-07	L	0	0.399109003	H
Nucleus Perimeter	1	1.60E-08	L	0	0.278611101	H
Nucleus Major	1	6.76E-09	L	0	0.136841267	H
Nucleus Angle	0	0.216745907	L	0	0.927478174	H
Nucleus Circularity	1	3.67561E-04	H	0	0.273993333	L
Nucleus Aspect Ratio	0	0.223888859	L	0	0.092011377	H
Nucleus Roundness	0	0.2625561	H	0	0.082731096	L
Nucleus Solidity	0	0.074281351	H	0	0.124702902	L
	SL v SH on SET			**ML vs MH on SET**		
	Significant (1) or not (0)	p-value	Cell Line with higher mean	Significant (1) or not (0)	p-value	Cell Line with higher mean
Area	1	1.64641E-04	L	1	1.49E-17	H
Perimeter	1	1.15E-06	L	1.00E-00	3.21E-21	H
Major axis	0	0.217680048	0	1	7.31E-14	H
Minor axis	1	1.84E-10	L	1	3.38E-20	H
Circularity	1	3.24E-05	H	1	1.94E-15	L
Aspect Ratio	1	2.12E-06	H	0	0.875889932	L
Roundness	1	1.26E-05	L	1	9.62E-06	L
Solidity	1	0.00673473	H	1	3.05E-24	L
Max Span Hull	0	0.976305895	0	1	1.29E-15	H
Area Hull	1	7.10E-05	L	1	8.41E-19	H

Perimeter Hull	0	0.168120354	0	1	1.29E-20	H
Circularity Hull	1	6.66E-05	L	0	0.111580393	L
Max Rad Hull	0	0.741891225	0	1	1.86E-15	H
Max/Min Rad Hull	1	4.48139E-04	H	1	0.048981042	L
CV Rad Hull	1	0.002198444	H	1	0.014724811	L
Mean Rad Hull	0	0.747565824	0	1	2.13E-19	H
Diameter Bounding Circle	0	0.927051607	0	1	6.85E-16	H
Max Rad Circle	0	0.925962896	0	1	6.85E-16	H
Max/Min Circle	1	0.041942415	H	1	0.0064255	L
CV Rad Circle	1	0.006119484	H	1	4.95E-06	L
Mean Rad Circle	0	0.715080747	0	1	1.84E-19	H
Nucleus Area	1	1.91E-07	L	0	0.889749117	H
Nucleus Perimeter	1	7.18E-07	L	0	0.20859413	L
Nucleus Major	1	0.005380744	L	1	0.004868009	L
Nucleus Angle	1	3.16E-10	L	0	0.825562251	H
Nucleus Circularity	1	0.002951905	H	1	6.79E-14	H
Nucleus Aspect Ratio	1	8.20E-06	H	1	2.97E-08	L
Nucleus Roundness	1	2.01E-05	L	1	1.65E-08	H
Nucleus Solidity	1	1.64E-05	H	1	4.05E-09	H

Table S5.3: True positive and true negative rates of different thresholds for sample classification.

Here we show the true positive and true negative rate data against different thresholds used to classify a sample of cells as belonging to the high-met or low met lines. Thresholds tested were [0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]. In each case, if the proportion of cells in the sample of class high-met were greater than the threshold, the sample was classified as belonging to that class. Good predictions are considered to be those where both true positives and true negatives were equal to or greater than 0.8 (i.e. 80%). It is clear that a threshold of 0.6 has the best performance. The abbreviations used to describe cell lines and the surfaces are as follows: GDA: Glass Detergent washed; GAA: Glass Acid etched; SET: Siliconized glass, Ethanol Treated. L: low metastatic cell line; H: high metastatic cell line. The cell lines are referred to as follows: (i) DL: DUNN; (ii) DH: DLM8; (iii) KL: K12; (iv) KH: K7M2; (v) SL: SAOS2; (vi) SH: LM7-Saos; (vii) ML: MG63; (viii) MH: MG63.2. TPR stands for true positive rate and TNR stands for true negative rate.

Comparison	Classes	Threshold	0.3	0.4	0.5	0.6	0.7	0.8	0.9
D Lines GAA									
	DH	TPR	1	1	0.99	0.93	0.79	0.41	0.12
	DL	TNR	1	1	1	1	1	1	1
D Lines GDA									
	DH	TPR	1	1	0.99	0.84	0.49	0.24	0.06
	DL	TNR	0.06	0.22	0.51	0.79	0.97	1	1
D Lines SET									
	DH	TPR	1	1	1	0.93	0.67	0.25	0.09
	DL	TNR	0.72	0.96	1	1	1	1	1
D Lines All surfaces									
	DH	TPR	1	1	1	0.94	0.87	0.6	0.21
	DL	TNR	0.14	0.38	0.64	0.84	0.92	1	1
K Lines GAA									
	KH	TPR	1	1	1	0.99	0.98	0.82	0.39
	KL	TNR	0.38	0.73	0.97	1	1	1	1
K Lines GDA									
	KH	TPR	1	1	1	1	1	0.98	0.68
	KL	TNR	0.38	0.73	0.95	0.99	1	1	1
K Lines SET									
	KH	TPR	1	1	0.97	0.85	0.57	0.32	0.05
	KL	TNR	0	0.03	0.16	0.47	0.82	0.95	1
K Lines All surfaces									
	KH	TPR	1	1	0.97	0.82	0.51	0.27	0.08
	KL	TNR	0.25	0.52	0.82	0.92	0.99	1	1

M Lines GAA									
	MH	TPR	1	1	1	1	0.94	0.71	0.32
	ML	TNR	0.75	0.97	1	1	1	1	1
M Lines GDA									
	MH	TPR	1	1	1	1	1	0.99	0.7
	ML	TNR	0.97	1	1	1	1	1	1
M Lines SET									
	MH	TPR	1	1	1	1	1	0.96	0.75
	ML	TNR	0.93	1	1	1	1	1	1
M Lines All surfaces									
	MH	TPR	1	1	1	0.97	0.88	0.7	0.41
	ML	TNR	0.83	0.94	0.97	0.99	1	1	1
S Lines GAA									
	SH	TPR	1	1	1	1	1	0.95	0.6
	SL	TNR	0.04	0.25	0.53	0.8	0.98	1	1
S Lines GDA									
	SH	TPR	1	1	1	0.96	0.82	0.51	0.2
	SL	TNR	0.27	0.62	0.89	0.99	1	1	1
S Lines SET									
	SH	TPR	1	1	1	0.96	0.81	0.5	0.14
	SL	TNR	0.11	0.34	0.62	0.89	1	1	1
S Lines All surfaces									
	SH	TPR	1	1	0.98	0.93	0.77	0.45	0.22
	SL	TNR	0.12	0.33	0.56	0.83	0.96	1	1
All H vs L GAA									
	H	TPR	1	0.99	0.98	0.9	0.65	0.38	0.1
	L	TNR	0.04	0.18	0.39	0.66	0.86	0.94	0.97
All H vs L GDA									
	H	TPR	1	1	0.99	0.96	0.77	0.48	0.27
	L	TNR	0	0.05	0.17	0.37	0.7	0.92	0.99
All H vs L SET									
	H	TPR	1	0.99	0.92	0.75	0.53	0.31	0.12
	L	TNR	0.1	0.35	0.64	0.84	0.97	1	1
All H vs L All surfaces									
	H	TPR	1	1	0.98	0.92	0.73	0.42	0.16
	L	TNR	0.11	0.33	0.48	0.68	0.93	1	1

Type-1 H vs L GAA									
	H	TPR	1	1	0.98	0.93	0.78	0.45	0.23
	L	TNR	0.07	0.21	0.49	0.73	0.9	0.98	1
Type-1 H vs L GDA									
	H	TPR	1	1	1	0.86	0.7	0.46	0.22
	L	TNR	0.24	0.49	0.75	0.95	0.98	1	1
Type-1 H vs L SET									
	H	TPR	1	1	1	0.95	0.88	0.56	0.29
	L	TNR	0.28	0.57	0.78	0.97	1	1	1
Type-1 H vs L All surfaces									
	H	TPR	1	0.98	0.94	0.88	0.69	0.38	0.17
	L	TNR	0.32	0.54	0.72	0.89	0.99	1	1